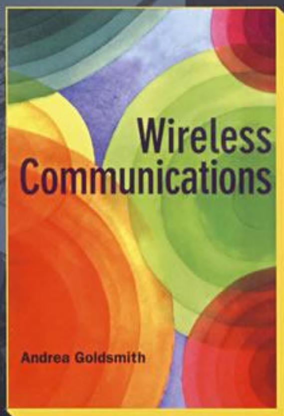


图灵电子与电气工程丛书

无线通信

Wireless Communications

[美] Andrea Goldsmith 著
杨鸿文 李卫东 郭文彬 等译





人民邮电出版社
POSTS & TELECOM PRESS

图书在版编目（CIP）数据

无线通信 /（美）哥德史密斯，（Goldsmith, A.）著；杨鸿文等译.
—北京：人民邮电出版社，2007.6
（图灵电子与电气工程丛书）

ISBN 978-7-115-15993-9

Ⅰ. 无… Ⅱ. ①哥…②杨… Ⅲ. 无线电通信 Ⅳ. TN92

中国版本图书馆 CIP 数据核字（2007）第 039503 号

内容提要

本书重点论述当前各类无线通信系统中具有普遍性和代表性的基本知识，包括基本的理论、问题、设计思路和分析方法。全书内容包括无线信道模型、无线信道容量、无线通信中的调制编码技术及均衡处理技术、扩频通信，还包括多天线系统、多用户系统、多载波调制、自适应调制与编码、蜂窝系统及无线自组织网络等。书中每部分内容的讲述都有全新的视角和独特的处理方法，并配合丰富的图示、例题和习题。

本书适合作为通信工程和电子信息类相关专业高年级本科生和研究生的教材，同时也可供工程技术人员参考。

图灵电子与电气工程丛书

无 线 通 信

◆ 著　　[美] Andrea Goldsmith
　译　　杨鸿文　李卫东　郭文彬　等
　责任编辑　朱　巍

◆ 人民邮电出版社出版发行　北京市崇文区夕照寺街 14 号
　邮编　100061　电子函件　315@ptpress.com.cn
　网址　http://www.ptpress.com.cn
　北京铭成印刷有限公司印刷
　新华书店总店北京发行所经销

◆ 开本：700 × 1000　1/16
　印张：32.75
　字数：739 千字　　2007 年 6 月第 1 版
　印数：1～4 000 册　　2007 年 6 月北京第 1 次印刷

著作权合同登记号　图字：01-2006-0312 号

ISBN 978-7-115-15993-9/TN

定价：69.00 元

读者服务热线：(010)88593802　印装质量热线：(010)67129223

前　　言

无线通信是一个广阔而又充满生机的领域，在过去的几十年里涌现出无数令人兴奋的技术成果。本书的目的是使读者对无线通信的基本原理有广泛的理解。这些原理包括无线系统的特性和性能极限，对其进行分析的方法和数学工具，还有与无线系统设计有关的深入理解和利弊权衡。同时用各种现实的和预想中的无线通信系统来引出和例证这些基本原理。本书可以用于无线通信领域高年级或研究生水平的教学，也可以作为该领域技术人员和科研人员的参考书。

本书的结构

第1章首先对无线通信进行综述，包括对历史的回顾和对未来的展望，以及对现有系统和标准的概览。第2章和第3章讲述了无线信道的特性，正是这些特性引出了无线系统设计中的很多挑战。其中，第2章的内容涵盖了无线信道中的路径损耗和阴影衰落特性，它们经过较长的距离才发生变化。第3章描述了多径导致的平衰落和频率选择性衰落，它们的变化距离短得多，只有信号波长数量级。第4章讨论无线信道的容量极限及达到该极限的传输方式。虽然在研究这些传输方式时没有考虑复杂度和时延问题，但它们同样对一些获得实际应用的技术具有重要意义，这些技术将在后面章节陆续介绍。第5章和第6章主要介绍无线通信中的数字调制技术及其性能。从中可见，衰落会明显降低性能，所以抗衰落技术对高性能无线通信系统是至关重要的。

接下来的一些章节介绍了减轻平衰落和频率选择性衰落的主要技术。其中，第7章介绍了分集技术的基本原理，包括一种可以明显简化性能分析的新数学工具。这些技术能够去除平衰落的大部分不利影响。第8章全面介绍了各种编码技术，包括已经成熟的分组编码、卷积编码和格码，以及级联码、Turbo码、LDPC码方面的新进展。本章表明，虽然噪声信道编码技术几乎接近了理想性能，但无线系统编码的设计和性能分析方面依然存在许多待解决问题。第9章研究了平衰落中的自适应调制技术，它利用无线信道的时变特性使通信变得健壮且具有较高频谱效率。本章也将自适应调制技术及性能与平衰落信道的容量极限联系在一起。第10章介绍了多天线技术和空时通信系统，空间维的引入带来了高数据速率和对衰落的健壮性。第11章介绍了均衡，它的作用是在接收机中进行信号处理以补偿频率选择性衰落。第12章介绍多载波调制，它在抗频率选择性衰落方面比均衡更简单灵活。第13章介绍了单用户和多用户的扩频技术。这些技术不仅可以减轻频率选择性衰落的影响，还允许多个用户共享无线频谱。

本书的最后3章着重介绍多用户系统和网络。第14章介绍了用于多用户共享无线信道的多址接入和随机接入技术，无论它们的数据流是连续的还是突发的。本章还涵盖了功率控制这种旨在减小多用户间干扰且确保各用户达到预期性能的机制。本章以多用户信道的容量极限及达到此极限的传输策略和信道共享技术结束。第15章涵盖了蜂窝系统的设计、优化和性能分析，以及与此类系统中功率控制和性能极限相关的高级主题。最

后，第16章介绍了自组织无线网络的基本原理和待解决问题。

背景知识

阅读本书的唯一要求是读者需要对概率论、随机过程及傅里叶分析有基本了解。如果有数字通信的背景会很有帮助，但本书也讲述了其中的基本原理，所以不是必需的。本书用三个附录概括了各章所需的主要背景知识。其中，附录A讨论了带通信号和系统的等效低通模型，以简化带通系统的分析。附录B概述了全书涉及的概率论和随机过程的主要概念。附录C介绍了矩阵相关的定义、结论和性质，第10章和第12章运用了这些知识。附录D总结了现有无线系统的主要特性及其标准。

本书特点

由于人们在无线领域中已进行了大量研究，加之无线系统设计很复杂，我们要对本书涵盖的所有问题都进行详尽的介绍是不可能的。所以每章最后都提供大量的参考文献，其中的文献对该章内容有更详细的介绍。本书还包含了近100个实例来例证和强调关键的原理和权衡。另外，本书还收录了约300道课后习题。这些题目分为几大类，用来复习和巩固正文中的内容。其中一些习题旨在对一些关键概念进行例证或进一步深化，同时给出或说明基于这些概念的无线系统的属性；有些习题用来证明正文中给出但并未推导的结论；还有一些习题对典型无线系统中的参数和性能给出了数值解，有些习题还引入了书中未涉及的新概念或系统设计。

教材使用

可以针对课程长短、学生背景情况及课程侧重点将本书作为教材灵活使用。本书的核心部分在第1章至第6章。此后的每章都涵盖一个单独的专题，这些章节在授课中可以省略或在其他课程中介绍。学习本书需要有信号与系统（包括模拟与数字）和概率理论与随机过程的本科课程基础。如果学生有数字通信方面的先修或并修课程，也会有所裨益，这样的话第5章的内容及其他章节中重复的内容可以作为复习内容而快速学习。

本书可以分为三个部分：第1章至第6章为核心内容，第7章至第13章为单用户无线系统设计，第14章至第16章为多用户无线网络。书中大部分内容可以分2至3个小学期或2个完整学期讲授。如果分3个小学期授课，则可以按照本书所分的3个部分学习并在最后做一个深入的研究课题。如果分2个完整学期，第一阶段可以学习第1章至第10章（单用户平衰落），第二阶段学习第11章至第16章（频率选择性衰落、多用户系统和无线网络）。对于只有1个小学期或完整学期课程安排的情况，重点可以放在单用户无线系统部分，即第1章至第6章及第7章至第13章中的部分章节。如此，可以通过后续学期的选修课程讲述多用户系统和无线网络（第13章的部分内容和第14章至第16章）。我在斯坦福大学授课时使用了两学期的分法，其中第2学期每隔一年开办一次，包括一些附加文献资料和一些深入的研究。一个学期的课程也可以同时包含第1章至第6章的单用户系统和第13章至第16章的多用户系统，如果时间允许的话，还可以加入第7章至第12章中的一些专题。

配套网站（http://www.cambridge.org/9780521837163）提供了本书的补充资料，包括授课用的幻灯片、附加练习和勘误表。

致谢

完成一本书需要耗费大量的人力，我对在本书编写过程中所有帮助过我的朋友深表谢意。首先要感谢加州理工学院和斯坦福大学使用本书各版本初稿的10届学生，他们的建议、见识和经验对本书主题的选择、内容涵盖范围及语言语气都有很大价值。John Proakis和其他几位匿名审稿人对本书的初稿提出了有价值、有深度的意见和建议，并指出了其中的遗漏和不足，这大大提高了最终稿的质量。我的在读研究生Rajiv Agrawal、Shuguang Cui、Yifan Liang、Xiangheng Liu、Chris Ng和Taesang Yoo非常认真地校对了初稿的许多章节，提出了新的观点，重新推导了公式，检查了打印稿并指出了其中的错误和遗漏。已毕业的学生Tim Holliday、Syed Jafar、Nihar Jindal、Neelesh Mehta、Stavros Toumpis和Sriram Vishwanath仔细地检查了各章并提出了有价值的建议。另外，我所有在读的和已毕业的学生（包括那些已经提过的和Mohamed-Slim Alouini、Soon-Ghee Chua、Lifang Li和Kevin Yu）的许多研究结论对本书的许多内容都是有力的支持，尤其在第4章、第7章、第9章、第10章、第14章和第16章中。习题解答是由Rajiv Agrawal、Grace Gao和Ankit Kumar编写的。我同样要感谢我的很多同事，他们在百忙中抽出时间来阅读并指正某些章节。他们的批评指正极为亲切、慷慨和诚恳。他们深邃而有价值的观点不仅极大地完善了本书，而且也让我在无线方面受益匪浅。我还要特别感谢Jeff Andrews、Tony Ephremides、Mikė Fitz、Dennis Goeckel、Larry Greenstein、Ralf Koetter、P.R.Kumar、Muriel Médard、Larry Milstein、Sergio Servetto、Sergio Verdú和Roy Yates为本书所做的贡献。Don Cox无私奉献着他在工程方面的智慧，在无线系统的许多细节和设想上给了我很大启示。我也非常感谢多年与我共同进行研究的研究人员以及在Maxim 技术公司和AT&T贝尔实验室的同事们，他们开阔了我在无线通信和相关领域的视野。

我非常感谢斯坦福大学的同事、学生和领导们，他们使本书诞生于充满活力且令人兴奋的教学和科研环境之中。我也很感激美国海军研究办公室和美国国家自然科学基金会在本书的整个编写过程中给予的资金支持。还要感谢我的助理Joice DeBolt和Pat Oshiro，他们为了支持我的研究和教学，事无巨细地帮助我。我还想感谢加工编辑Matt Darnell在整个出版过程中对一些细节的把握。本书策划编辑Phil Meyler从10年前这本书的策划到现在都一直关注着它。他对此书的鼓励和热情始终如一，并宽容地接受了本书编写过程中出现的变化和延期。我想不会有更好的编辑能承担如此一项有所回报但又异常艰难、繁重的任务了。

我还要特别感谢两个人，他们自始至终支持着本书和我的事业。Larry Greenstein通过他渊博的知识和科研经历，点燃了我最初对无线的兴趣，他是我的知识源泉和良师益友。Pravin Varaiya作为我的博士导师和我的榜样，他渊博的知识和惊人的严谨见识与热情深深地影响着我。他是我的动力和灵感之源。

对于我的朋友和家人给予的关爱、支持和动力，我也深表感激。他们为我提供了一

个坚实的支撑网络，如果没有他们，这本书将不会完成。我要特别地感谢Remy、Penny和Lili，他们给了我关爱和支持，感谢我的母亲Adrienne，她给了我爱并带给我创造力和写作的兴趣。我的父亲Werner深刻地影响着这本书和我的生涯。他是一名资深教授，一位成果颇丰的学者和作者，是机械工程和生物工程中许多领域的先驱[1]。他给我从事工程学的建议，使我的事业得以起步，他也是我最主要的啦啦队长。他的自豪、关爱和鼓舞一直支持着我。我有幸帮助他完成最后论文，在本书中也沿袭了与他合作时所感受到的严谨、对细节和拼写错误的一丝不苟。

最后，我无言以表达对丈夫Arturo和孩子Daniel、Nicole的爱和感激。Arturo对本书的编写和我事业的方方面面都给予了无限的支持，他为此做出了很多牺牲。他的鼓励、爱护和奉献，使我在经历学术和家庭生活的起起落落时能继续前进。他是我所梦想的最好的丈夫和朋友，他使我的生活更加充实。Daniel、Nicole是我的阳光，我每天都因他们的爱和甜蜜而更加灿烂。能与这三个特别的人共同生活，我感到无比幸运。谨以此书献给他们。

1. 作者的父亲Werner Goldsmith（1924—2003）是著名机械工程和生物工程学家，IEEE会士。他们父女均荣膺IEEE会士称号，在科技界一时传为佳话。——编者注

目　录

PDG

第1章 无线通信概述

无论从什么角度看，无线通信都是通信产业中发展最快的一部分。正因为如此，无线通信受到媒体的普遍关注，公众对它充满了期待。在过去的10年中，蜂窝系统经历了指数性的快速增长，全球用户数已达20亿。在大多数发达国家，蜂窝电话已成为工作和日常生活中不可缺少的工具。在许多发展中国家，蜂窝系统也在迅速取代过时的有线系统。许多家庭、商务区和校园已经开通了无线局域网，它正在补充甚至替代有线网络。无线传感器网络、自动化高速公路、自动化工厂、智能家居、智能家电、远程医疗等新的应用正从研究设想变为现实。无线通信爆炸式的增长以及笔记本电脑和掌上电脑的大量普及，充分显示出无线网络有着光明的前景。不过，要想设计性能足以支持这些新兴应用并且健壮的无线网络，我们还面临着许多技术上的挑战。本章首先简要回顾无线通信的历史：从前工业化时期的狼烟，到今天的蜂窝系统、卫星系统及其他无线通信网络。然后进一步详细讨论无线技术的愿景，包括有待解决的技术难题。本章还将介绍一些已有的或将要出现的无线通信系统和标准。这些系统和未来无线愿景的差异说明要实现这些无线愿景还有大量的工作要做。

1.1 无线通信的历史

最早的无线通信出现在前工业化时期，这些系统使用狼烟、火炬、闪光镜、信号弹或旗语，在视距（距离，后来因望远镜而有所增加）内传输信息。为了能传输更复杂的消息，人们又精心设计出了用这些原始信号组成的复杂信号。为了能传得更远，人们在山顶、道路旁建立了一些接力观测站。直到1838年，这些原始的通信网才被塞缪尔·莫尔斯（Samuel Morse）发明的电报网替代，接着又被电话取代。在电话发明几十年后的1895年，马可尼（Marconi）首次从英国怀特岛（Isle of Wight）到30km之外的一条拖船之间成功进行了无线传输，现代意义下的无线通信从此诞生。从这一天开始，无线通信技术迅速发展，使我们能够在更远的传输距离上实现更好的通信质量、更低的功耗、更小的体积和更便宜的价格，使公网和专网的无线通信、无线电视、无线网络等成为现实。

早期的无线通信传送的是模拟信号。今天大多数的无线通信系统传送的是由二进制比特组成的数字信号，这些比特或者直接来自于数据信号，或者是由模拟信号数字化得到。数字无线通信既可以连续发送比特流，也可以对比特进行分组后以数据包为单位进行发送。后一种情形称为*分组无线电*（packet radio），其特点是突发传输：数据包的发送可能是间歇性的，没有数据包发送的时候信道是空闲的。当然了，这不表示分组传输不能连续地传送数据包。第一个分组无线电网络是1971年夏威夷大学开发的ALOHANET。它把散布在4个岛上的7个校区的计算机节点通过无线传输连接到Oahu岛上的中央计算机。ALOHANET采用星型拓扑结构，中央计算机置于拓扑中心，网中的

任意两台计算机可通过中央计算机建立双向通信。分组无线电系统中的第一个信道接入协议和第一个路由协议就出现在ALOHANET中，其中的许多基本思想至今还在沿用。美国军方对ALOHANET把无线广播和分组数据相结合这一点产生了浓厚的兴趣。从20世纪70年代中期至80年代初，美国国防部高级研究计划署（Defense Advanced Research Projects Agency，DARPA）投入了巨大的人力物力来开发基于分组无线电的战术通信系统。在这些能够自组织的无线网络中，节点不需要借助任何基础设施就能够自己形成一个网络。DARPA对于自组织网络的投资在20世纪80年代中期达到顶峰，但是，所得到的系统在传输速率等方面的性能和预期的相去甚远。出于军事需要，这些网络目前仍然在继续研发中。分组无线电在商业方面的一个应用是支持广域无线数据业务。早在20世纪90年代初就引入了这种业务，它能提供较低速率（约20kbit/s）的无线数据接入，包括电子邮箱、文件传送和网站浏览。不过因为速率低、成本高以及缺乏“杀手级应用”等原因，这种广域无线数据业务自推出以来从未形成过强大的市场，在20世纪90年代基本消失，取而代之的是无线局域网，以及能同时支持数据的蜂窝电话网。

2

20世纪70年代，有线以太网技术的出现使许多厂商不再关心无线网络。以太网10Mbit/s的数据速率远远超过了当时的无线通信所能达到的速率。虽然有线网有敷设缆线的不便，但有了这样高的速率，公司也就不介意铺几根线了。一直到1985年，美国联邦通信委员会（Federal Communications Commission，FCC）授权无线局域网产品可以开放使用工业、科学和医疗（Industrial Scientific and Medical，ISM）频段，这才启动了无线局域网的商业开发。无需FCC许可就能直接使用ISM频段，这对无线局域网商家来说是非常有吸引力的。不过为了不干扰原先分配于ISM频段的主要业务，FCC规定无线局域网必须要用较低的发射功率，并且只能用低频谱效率的信号方式。同时，无线局域网又要受到该频带内主要业务的强干扰。这些因素使最初的无线局域网在数据速率和覆盖范围方面的表现不尽如人意。再加上对安全性的顾虑、标准的不统一以及高成本（第一个无线局域网接入点的报价是1400美元，而有线以太网卡只是几百美元）等因素，导致早期的无线局域网产品销量很低。实际上，早期的无线局域网几乎没有用于数据网的，只能在库存处理这样的低技术应用中派上用场。目前使用的无线局域网采用IEEE 802.11系列标准，虽然数据速率也不算太高（最大速率为几十兆比特每秒），覆盖范围也比较小（大约100m），但性能上已有很大的提升。今天的有线以太网能提供1Gbps的数据速率，倘若没有额外的频谱分配，有线网和无线网的性能差距恐怕会越来越大。虽然无线局域网的数据速率相对较低，但由于其方便且不受连线束缚的优点，它正逐步成为许多家庭、办公室、校园等环境中互联网接入的首选。不过，大部分无线局域网目前支持的业务，像电子邮件和网络浏览等，都还不是带宽密集型的，未来无线局域网所面临的一大挑战就是能够支持大量用户同时使用有时延约束的带宽密集型业务，如视频。同时扩大覆盖范围也是未来无线局域网发展的一个重要目标。

迄今为止，最成功的无线网络恐怕要数蜂窝电话网了。这类系统最早要追溯到1915年在纽约和旧金山之间建立的无线话音传输。到1946年，全美国有25个城市引入了公共移动电话业务。早期的这些系统采用一个中央发射机来覆盖整个城市。这种低效

的频率使用方式，加之当时的无线电技术水平也不高，使这些系统的容量都非常有限，以至于在引入移动电话业务的30年以后，纽约市的系统只能容纳543个用户。

20世纪五六十年代出现了解决容量问题的一种方案，这就是美国电报电话公司（AT&T）贝尔实验室的研究人员提出的蜂窝概念[1]。蜂窝系统基于这样一个事实，即传输信号的功率随着距离的增大而减小。把两个用户在空间上分隔足够远的距离，就能使它们之间的干扰非常小，于是这两个用户就可以使用相同的频率进行各自的通信，从而使频率被充分利用，系统所能承载的用户数也因此而大量增加。蜂窝系统从概念的提出到真正实现经历了一个漫长的过程。AT&T于1947年向FCC申请蜂窝电话业务的频段，并于20世纪60年代末基本完成了蜂窝系统的设计草案，但首次外场测试1978年才进行，FCC直到1982年才对蜂窝业务颁发许可，而在这个时候，许多初始的设计在技术上已经过时了。1983年第一个模拟蜂窝系统在美国芝加哥市开通，到1984年系统容量就已经饱和，为此FCC把分配给蜂窝系统的频段从40MHz增加到50MHz。蜂窝产业增长速度之快超出了所有人的预料。早在第一个蜂窝系统出现之前，AT&T曾经做过一份市场调查预测，认为需要蜂窝手机的只是医生等少数高收入人群。于是，AT&T在20世纪80年代基本上放弃了蜂窝网络这部分业务，把主要精力转向了光纤网络，后来到蜂窝网络逐渐显示出潜力的时候，AT&T才开始着手重新开展这项业务。到20世纪80年代后期，许多城市的蜂窝网络已经不能满足日益增长的需要。此时，开发具有更大容量和更好性能的数字蜂窝技术已经显得非常迫切。

3

基于数字通信技术的第二代蜂窝系统于20世纪90年代初面世。数字系统的高容量以及数字电路在成本、速度、硬件功率效率方面的优势带动了蜂窝系统从模拟到数字的转变。第二代蜂窝系统最初主要提供话音业务，后来逐渐演进到可以提供电子邮件、互联网接入、短消息等数据业务。不幸的是，蜂窝电话的巨大市场潜力导致了第二代蜂窝系统标准的多样性。仅在美国就有三个不同的标准，在欧洲和日本还有其他的标准，这些标准都是互不兼容的。不同的城市采用互不兼容的标准，这样一来，只用一部单标准手机在美国国内或世界范围内漫游就是不可能的。目前在一些国家正在着手进行的第三代蜂窝系统的开发中，也存在这种多种不兼容标准共存的现象。多模手机就是这种标准不统一的产物，这种手机集成了多种数字通信标准以便实现全国和全球漫游。考虑到只有模拟系统能够覆盖美国全境，也有一些多模手机同时集成了第一代的模拟通信标准。

卫星通信系统通常按其轨道的高度分为，高度在2000km左右的低轨道卫星系统（low-earth orbit，LEO）、高度在9000km左右的中轨道卫星系统（medium-earth orbit，MEO）以及高度约40000km的地球同步轨道卫星系统（geosynchronous orbit，GEO）。从地球看过去，GEO卫星是静止的，而其他轨道卫星的覆盖区域是随时间变化的。1945年，科幻小说作家Arthur C. Clarke最早提出了用GEO卫星进行通信的构想。不过第一颗人造卫星前苏联1957年发射的“伴星一号”以及美国国家航空航天局（NASA）和贝尔实验室于1960年发射的卫星“回声1号”（Echo-1）都不是GEO卫星，因为以当时的技术很难将卫星送到如此高的轨道。第一颗GEO卫星于1963年由美国休斯公司和NASA发射升空。在此后的几十年间，无论是在商用卫星通信系统中，还是在官方的卫

星系统中，GEO卫星一直占据着主导地位。

地球同步轨道卫星的覆盖区域非常大，只需要少量几颗卫星（即少量费用）就能覆盖广阔的地域，甚至可以覆盖全球。不过因为距离远，所以需要很大的发射功率才能使信号到达卫星，并且对于像话音这样有时延约束的业务而言，传播时延也显得太大。因为这些缺点，人们在20世纪90年代将兴趣转向了低轨道卫星[2~3]，期望建立一个话音和数据业务能和地面蜂窝系统竞争的卫星系统。但卫星通信系统的终端与蜂窝电话相比，毕竟个头大、耗电量大、成本高，这些因素大大影响了卫星系统的普及。卫星通信的最大优势是它的覆盖能力，特别适合于偏僻地区和有线或蜂窝电话尚未充分铺开的第三世界国家。但这些地区对通信的需求一般不大，居民一般也没有能力支付卫星通信的费用。而LEO系统在人口密集地区本来可以获得的收入中的绝大部分也被越来越普及的蜂窝系统夺走。由于没有市场，绝大多数的低轨道卫星系统逐步淡出了商业领域。

卫星通信最适合的领域是广播娱乐节目。工作在12GHz频段的直播卫星可以发送几百路的电视节目，足以和有线电视竞争。在欧洲和美国运行的卫星数字广播也逐步开始流行，这些系统能提供接近CD音质的数字音频广播。

1.2 无线愿景

展望未来，无线通信在几十年内仍将是通信发展的前沿。无线通信的愿景是，不论在世界的任何地方，人们都可以用小体积的手持设备或笔记本电脑方便地进行多媒体通信。不论是在办公室、校园，还是街边的咖啡厅，掌上电脑、笔记本电脑和台式电脑都可以通过无线网络互相通信。在居室里，无线网络不仅能连接电脑、电话和安全监视系统，还可以使新型的智能家电交互式工作，甚至能通过互联网控制这些家电。对于老弱病残，这种智能家居可以照料他们的生活起居，可以帮助监视病人，并在紧要时刻做出应急响应。无线娱乐服务也将渗透到家庭和其他公共活动场所。视频电话会议可以将彼此分隔的人们联系在一起，不论是相隔几个街区还是远隔重洋，错过航班的销售人员或是在加勒比海度假的CEO也照样可以加入到会议中来。无线视频使得远程教育、远程培训和远程医疗成为可能。无线传感器在民用和军用方面都有巨大的价值。民用用途包括火灾监控、有害废弃物处理、建筑物和桥梁的受力监测、二氧化碳气体运动和受灾地区有害气体扩散的监控等方面。这些无线传感器能自行构成网络，对传感器的测量结果进行处理后，将信息发送给中央控制器。军用用途包括识别和跟踪敌方目标、探测生化袭击、辅助无人驾驶车辆以及反恐活动等。无线网络还能使分布式控制系统中的远程设备、传感器和制动器通过无线信道相连接，这样的系统能使得自动化高速公路、移动机器人和可重构的工业自动化成为现实。

以上所描述的多种应用都是无线技术愿景的组成部分。那么，无线通信准确地说是什么呢？这是一个复杂的话题，存在许多的方式可以把这个问题分解到不同层面上，包括不同的应用、不同的通信系统或是不同的覆盖区域[4]。无线应用包括话音、互联网接入、网页浏览、寻呼及短信、用户信息服务、文件传送、视频电话会议、娱乐、传感器网络以及分布式控制系统。无线通信系统包括蜂窝电话系统、无线局域网、广域无线数

据网、卫星通信网以及无线自组织网。无线覆盖区域包括室内、校园、城市、地区及全球。对于不同的分类方式，无线通信的表现特征不同，导致了无线通信产业界的丰富细化，体现为大量不同的无线通信产品、标准以及已经提供或拟议提供的业务。出现这种多样性的一个原因是，不同的无线应用有不同的要求。话音系统需要数据速率相对较低（约20kbit/s），能允许较高的误比特率（bit error rate，BER）（约10^{-3}），但总时延必须小于100ms，否则对端用户就能觉察到这个时延[1]；典型的数据系统则要求很高的数据速率（1Mbit/s～100Mbit/s）和非常低的误比特率（10^{-8}或更小，并且所有出错的比特都必须要重传），但不要求时延固定；实时视频系统需要的数据速率很高，对时延的要求和话音系统相当；寻呼和短信需要的数据速率很低，没有硬性的时延要求。需求的多样性使我们很难建造一个单一的无线通信系统，它能同时有效地满足所有要求。而这样的目标对有线网络是可能的，因为有线网络的数据速率能达到吉比特每秒，误比特率可低至10^{-12}。而无线网络因为数据速率低，误比特率又高，所以至少在不久的将来，无线系统依然会保持这种细分的形势，不同的应用有各自不同的协议以满足不同的要求。

5

蜂窝电话和无线互联网接入的快速增长使人们对无线通信的未来普遍感到乐观。不过，并不是所有的无线通信应用都是十分成功的。在很多无线通信系统享受着巨大成功的同时，也有许多系统失败了，其中包括第一代无线局域网、铱星系统、包括Metricom在内的广域数据业务以及家庭固定无线接入（无线“电缆”）等。谁也不能预言将来哪一种网络会成功，哪一种会失败。最为重要的是，无线通信的工程师和管理部门必须要有充分的灵活性和足够的创造性，这样才能抓住稍纵即逝的成功机遇。不过，从今天已有的和正在出现的各种无线通信网络中，从无线通信能够达到的应用愿景中，我们可以肯定，无线通信的未来必定会一片光明。

1.3 技术问题

为了使未来的无线应用得以实现，首先必须要攻克许多技术上的难题，这些难题贯穿于无线通信系统设计的各个方面。随着无线终端的功能越来越多，这些小巧的设备必须兼容多种工作模式、支持多种应用、适应多种信息媒体。我们知道，电脑能够处理语音、图像、文本、视频等不同性质的数据。要让价格便宜、轻便的手持设备也能做到这些，电路设计方面就必须要有新的突破。消费者不愿意使用需要经常充电的笨重电池，因此便携终端在信号发送及处理方面消耗的能量一定要最小。处理多媒体和实现网络功能所需要地信号处理过程一般比较耗电，因此在有基础设施的架构式无线网络，例如无线局域网和蜂窝系统中，要尽可能多地将信号处理负荷安排在能量充足的固定站点上。与此同时，架构式网络必须要重视可能出现的处理瓶颈、单点故障等问题。无基础设施架构的无线自组织网络有灵活、健壮的特点，适合于多种应用。这些网络中的各种处理和控制都必须在各网络节点上分布式执行，不存在一个能量充足的固定站点，因此如何有效利用能源是一个难题。在一些节点无法充电的网络，比如无线传感网络，能量成为

1. 有线电话的时延限制约为30ms，蜂窝电话将此限制放宽到100ms，IP电话又有所放宽。

一个特别重要的资源。对于这一类对能量有硬性限制的应用，如何优化网络设计是技术上的一个挑战。另外，无线信道不仅带宽有限，而且有随机多变的特性，网络难以保持固定的性能，这也要求应用足够健壮。

6

无线信道的特性决定了无线网络设计与有线网络设计的截然不同。随机的无线信道不是理想的传输媒介。首先，无线频谱是稀缺资源，必须分配给不同的系统和业务使用，因此无线电频谱必须由区域性和全球性的管理机构控制。工作于给定频段的区域性或全球性无线通信系统必须遵守相应管理机构对这一频段做出的种种规定。频谱还有可能是非常昂贵的，许多国家的频谱许可证通常颁发给竞标中出价最高的买家。在美国，运营商花费了90亿美元才获得第二代蜂窝系统的运营许可。在欧洲，第三代蜂窝系统运营许可的拍卖收入高达1000亿美元。通过这样高价竞拍得到的频谱必然要求非常充分地利用以获得投资回报，这意味着频谱将被高度复用，而且要求蜂窝系统的设计必须满足大容量和高性能的要求。在几吉赫兹的频段上，无线通信器件容易做到大小合适、功耗适中、成本低廉。但这一频段已经拥挤不堪，因此如果技术上的突破能使系统可以在更高的频段工作而成本和性能不变，就能大大缓解频谱资源的不足。不过在高频段工作时，如果采用全向天线，路径损耗将增大，传输距离将受到限制。

当信号在无线信道传播时，如果发射机、接收机或周围物体在运动，多径反射和衰减的变化将使信号经历随机波动。因此，无线信道的特性是不确定的、随机变化的，这就使我们难以设计出一个可靠并且性能可以保证的系统。无线通信的安全性也是一个很难解决的问题，任何人通过一副射频天线就可以轻松地截获电波。模拟蜂窝系统没有安全保证，只要扫描模拟蜂窝系统所在的频段就能容易地窃听到用户的谈话。数字蜂窝系统都采取了一定的加密措施。但如果有足够的技术、时间和决心，大部分加密方式都能被破解，而且有些确实已经被破解了。为了支持电子商务、信用卡交易这样的业务，无线网络的安全性必须进一步加强。

无线组网也是一个巨大的难题。网络必须能够从遍布全球的数十亿移动终端中确定出特定用户的位置。即使用户以100km/h的速度运动，网络也必须能够接通拨打给这个用户的电话。用户的位置和需求是变化的，与之相适应，网络的有限资源也必须以一种公平有效的方式进行分配。此外，目前还存在着拥有庞大基础设施的有线网络：电话网、互联网以及光纤网，我们可以把这些网络同无线网络联系在一起从而构成一个全球网络。然而在数据传输速率和可靠性方面，无线网络永远无法与有线网络相媲美。无线网络和有线网络接口之间巨大的性能差异也是一个棘手的问题。

7

无线网络设计的最大技术难题也许就在于设计流程本身的变革。有线网络大都依据分层原则进行设计，各层协议独立运作，层与层之间通过接口连接。无线通信系统中有这样一些层：物理或链路层负责媒质内的比特传输，接入层负责传输媒质的共享接入，网络及传输层负责网络中的路由及端到端的连接和数据传送，应用层根据具体的应用规定端到端的数据速率和时延。分层方法虽然降低了复杂性且易于模块化和标准化，但只能做到每层优化，不是全局优化，这使得效率和性能都会差一些。在有线网络中，分层设计虽然对话音、视频这类有时延约束的业务来说性能上会有所限制，但总的来说，网

络的高容量和高可靠性掩盖了分层设计的负面因素。无线网络的情形与之完全不同。无线链路的性能差，且性能及用户的连接状况、网络的拓扑结构都在随时间变化。无线链路的概念自身也相当模糊，它不像有线链路那样是一根具体的连线，而是体现为无线电波的传播、广播特性。无线信道的这些特性说明高性能的无线网络设计必须要针对信道进行优化，使其对信道或网络的动态变化有健壮性。因此，无线网络从链路层到应用层的每一层协议都需要高度的配合并具有自适应的能力。这样的跨层协议设计需要包括通信、信号处理、网络等在内的多学科知识。

下一节将对目前正在运行的各种无线系统进行概述。我们将看到，现有系统离无线通信的愿景还有一定的距离，尚有许多关键技术问题需要解决。这些技术问题就是本书的内容。

1.4 现有无线系统

本节将简单回顾目前正在运营的无线通信系统。这些系统的具体设计还在不断演进中，不断有新的系统出现、旧的系统退出。本节只关注大多数常见系统中的高层次设计问题，关于无线标准方面的具体资料可参阅文献[5~7]。本书附录D是主要无线标准的小结。

1.4.1 蜂窝电话系统

蜂窝电话系统在世界范围内普及并获利丰厚，它激发了无线通信技术的革命性进步。蜂窝系统可提供双向话音和数据通信服务，服务区域可以是一个地区、一个国家或者是全世界。蜂窝系统最初是为车载移动终端设计的，其天线置于车辆的顶部。现在已发展到可支持轻便的手持终端，用户可以在室内或室外使用，可以在步行或乘车时使用。

8

蜂窝系统的基本原理是频率复用，它利用信号功率随传播距离增大而减小的特性，允许空间上分开一定距离的两个点使用同样的频率。蜂窝系统将覆盖区域划分成许多互不重叠的小区，每个小区使用一组信道。同一组信道可被一定距离之外的另一小区重复使用，如图1-1所示，C_i表示在该小区所使用的一组信道。位于小区中央的基站控制着小区内的各种操作，下文将详细阐述复用相同信道组的不同小区之间的干扰称为小区间干扰（intercell interference)。复用相同信道组的小区之间的距离称为*复用距离*(reuse distance)。复用距离应尽可能小以增加频率复用的次数，从而最大地提高频带利用率。但复用距离减小时，同频小区间的距离也减小，干扰就会增大。为保证系统性能，小区间干扰必须低于某个给定的门限值，因此复用距离也不能低于某个最小值。因为无线信号的传播特性会造成有用信号和干扰信号的功率随机变化，因此在实际系统中确定这个最小的复用距离并非易事。为了最佳地决定复用距离和基站位置，必须对小区中的信号传播情况作出精确的描述。

早期的蜂窝系统设计中，因为基站设备相当昂贵（单价约100万美元），所以系统采用较少的小区个数来覆盖整个城市或地区。基站通常安装在高大建筑物上或山顶上，基站的发送功率非常大，小区的覆盖范围可以达到几平方英里，这种小区称为*宏小区*

(macrocell)。基站的信号沿各个方向均匀辐射，若无其他物体阻挡，移动台在基站周围做圆周运动时接收到的功率基本不变。这种圆形的等功率线交割的结果形成了正六边形的小区形状。在能够互不重叠同时又能完整覆盖整个服务区域的所有可能形状中，正六边形是最接近圆形的。

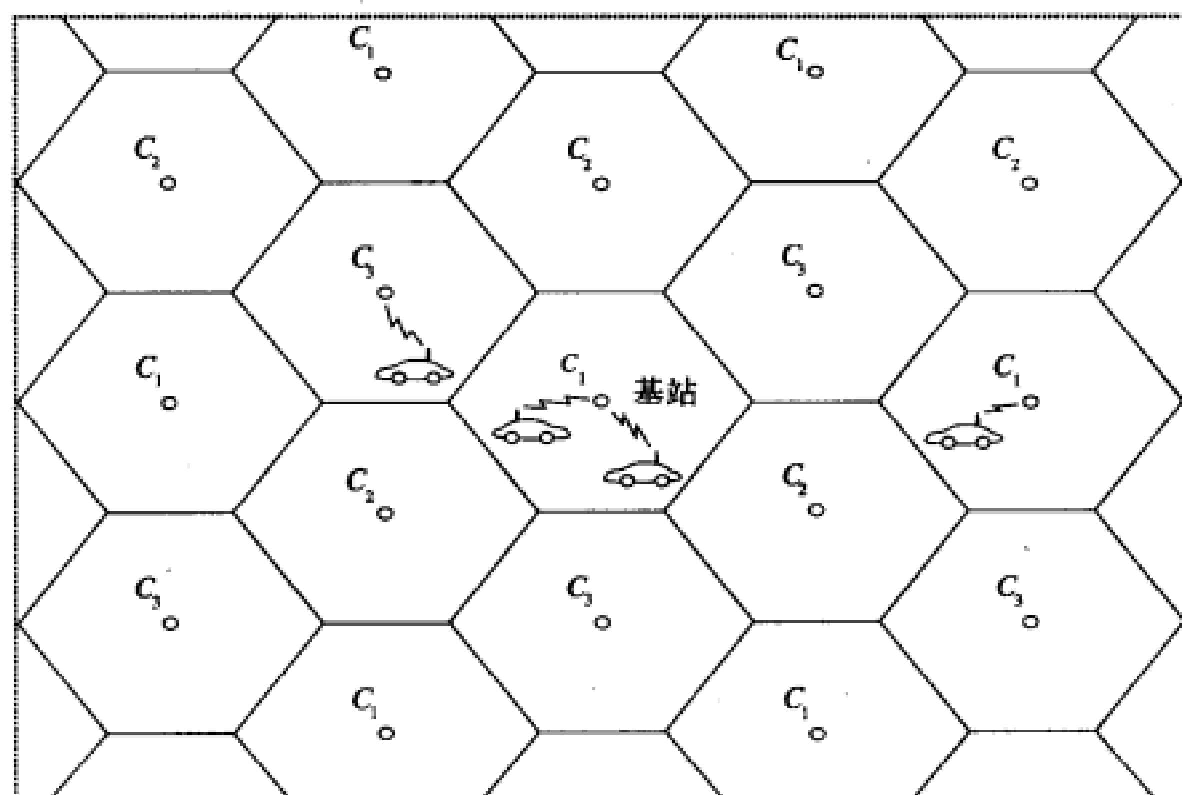


图1-1　蜂窝系统

城市地区现今的蜂窝系统大多采用很小的小区，基站邻近街道，发送功率也很低。视具体大小的不同，这些小区称为微小区或微微小区。促使这种演进的原因有两个：一是高密度用户地区需要有更大的容量，二是基站的体积和成本减小了。从系统的角度看，任意大小的小区可支持的用户数目相差不大，因此对于给定的覆盖范围，采用大量微小区时，单位面积上可支持的容量要显著高于只采用少量宏小区的情形。此外，采用微小区时终端和基站的距离小，所以需要的发送功率也小。不过微小区将使网络设计变得复杂化。移动台在微小区中的滞留时间短，因此必须要快速处理切换。位置管理也变得更复杂，因为用户所在地有很多个微小区。微小区情形下，建立通用传播模型也很困难，因为此时信号的传播高度依赖于基站的位置及周围反射体的几何分布。此外，在微小区传播条件下，正六边形不再很好地近似小区形状。微蜂窝系统设计时经常采用正方形或三角形的小区形状，这些形状在微小区信号传播模型下有很大的边缘误差[8]。

在图1-2中，在一定地理范围内的所有基站通过高速通信链路连接至移动电话交换局（mobile telephone switching office，MTSO）。MTSO相当于网络的中央控制器，它给每个小区分配信道，负责协调切换过程以及接续主呼和被呼电话。MTSO将电话接入到公众电话交换网（public switched telephone network，PSTN），还可以提供互联网接入。当某个用户想申请信道时，它先通过所在小区的控制信道发出请求信息，所在小区的基

站将这个请求转发给MTSO，如果该小区还有空闲信道，MTSO就会接受请求，否则请求将被拒绝。当基站或移动台发现接收功率正在减小，并接近某个最小门限值时，就会启动切换过程。基站通知MTSO移动台需要切换，MTSO发消息询问周围的基站是否能够探测到这个移动台的信号。如果有回应，MTSO将协调处理原基站和新基站间的切换过程。如果新基站无空闲信道可用，则切换失败，发生掉话。切换失败是蜂窝系统中发生掉话的原因之一，此外若移动台和基站之间的信号强度因信道的随机变化而低于门限，也会发生掉话。

10

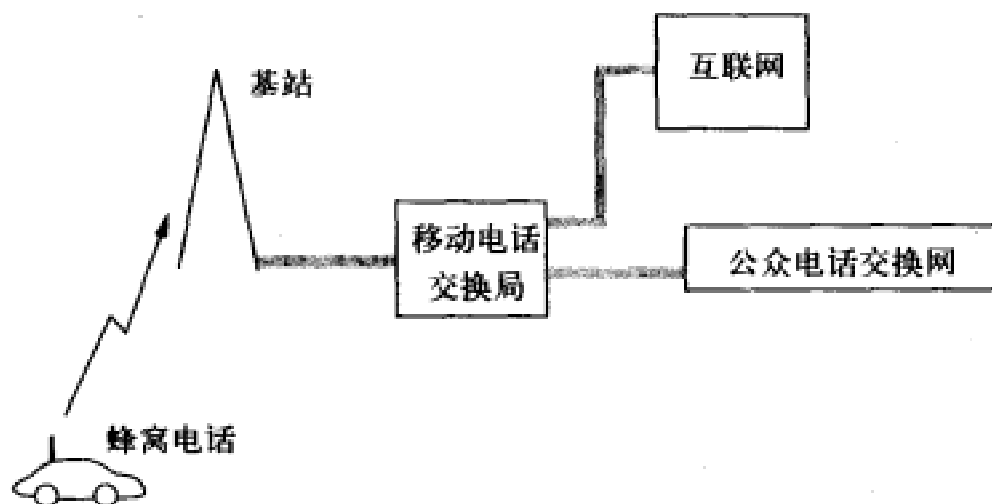


图1-2 现有蜂窝网络的结构

第一代蜂窝系统是模拟的，它们基本上是在数字通信还未普及的20世纪60年代所设计的。由于数字技术的诸多优点，第二代蜂窝系统转向数字设计。数字电路的器件价格便宜、处理速度快、体积小、耗电也少。数字通信中可以采用纠错编码及信号处理技术提高话音质量。依靠高频谱效率的数字调制技术及高效的频谱共享技术，数字系统可以有更大的容量。数字系统可以采用先进的压缩技术并能利用话音激活因子。数字系统可以应用加密技术来防止窃听。数字系统还能提供除话音之外的其他数据业务，包括短消息、电子邮件、互联网接入和图像服务（如可拍照手机）。由于数字系统低成本和高效率，服务供应商采取了有效的价格策略鼓励模拟用户转向数字系统。到今天，模拟系统主要用在还没有数字系统的地区。不过数字系统也非尽善尽美，有时存在话音质量差、频繁掉线、覆盖盲点等问题。随着技术和网络日渐成熟，系统的性能自然会稳步上升。有些地方的蜂窝系统所提供的话音质量和有线电话已不相上下。有些人甚至已经用蜂窝电话取代了家中的固定电话。

在通信系统中，频谱共享又称为多址接入，它是把信号空间沿着时间轴、频率轴或是码空间轴进行分解来实现的。频分多址（frequency-division multiple access，FDMA）把整个系统的带宽分割成若干个频率正交的信道。时分多址（time-division multiple access，TDMA）则将时间进行正交分割，每一信道在指定的时隙内占用整个频段。TDMA要求用户在时间上保持同步，所以实现起来要比FDMA困难一些。但TDMA易于支持多速率，只需给一个用户分配多个时隙即可。码分多址（code-division multiple access，CDMA）系统通常采用直序扩频或者跳频扩频方式，扩频码分为正交和非正交

两种情形。直序扩频把不同用户的数据调制到不同的码片序列上，码片序列的速率远高于数据速率。对应到频域，是窄带的数据信号与宽带的码片信号卷积，由此产生的信号带宽远远大于数据信号的带宽。在跳频中，调制器的载波频率受一个码片序列的控制，码片速率有可能比数据速率快，也可能比数据速率慢。结果形成了一个载波频率不断跳跃变化的跳频调制信号。不同用户的扩频信号通常在相同的带宽上互相叠加，扩频接收机用相应的扩频码将各用户的信号分离。但如果扩频码是非正交的，那么同一小区内的用户会产生小区内干扰（intracell interference），其他小区的扩频码也会带来小区间干扰（intercell interference）。扩频增益可以抑制这两种干扰。此外还可以通过多用户检测和干扰消除的手段进一步降低干扰。第13章和第14章将详细讨论不同的频谱共享技术及

11

其性能分析。如何选择多址技术是一个非常复杂的问题。对于给定的系统和运行环境，哪一种多址技术最佳一般不是可以简单回答的。

蜂窝系统是干扰受限（interference-limited）的，即系统中的干扰远远大于背景噪声，使得系统中用户数的增加主要受干扰程度的限制。因此，任何能够减小干扰的技术都能直接提高系统的容量和性能。此类已经应用或将要应用的技术包括小区扇区化、定向或智能天线、多用户检测、动态资源分配等，第15章将对此进行详细介绍。

美国的第一代蜂窝系统（1G）是AMPS（Advance Mobile Phone Service）。它采用频分多址，每个30kHz的信道传送一路FM调制的话音。起初FCC向该系统分配了40MHz的频谱，在该业务推广后没多久就增至50MHz以容纳更多的用户。系统的全部频带分为两个25MHz，一半用于移动台到基站的信道，另一半用于基站到移动台的信道。FCC又将这些信道分为两套，分别提供给每个城市的两个不同的运营商以鼓励竞争。欧洲与AMPS类似的系统是TACS（Total Access Communication System）。AMPS于20世纪80年代在全球范围内铺开，目前在一些地区包括美国的一些乡村，这仍然是唯一可用的蜂窝系统。

欧洲的第一代蜂窝系统中多数是互不兼容的。欧洲人很快达成了一致，为第二代数字蜂窝系统（2G）制定了一个统一的标准——GSM[1]。GSM结合了TDMA和慢跳频技术，话音调制采用频移键控。美国在制定2G蜂窝标准时，在多址技术上产生了激烈的争论，最终形成了多个互不兼容的标准[9-11]。具体来说，在900MHz的频段上有两个标准：IS-136[2]采用TDMA和FDMA的结合，调制方式是相移键控；IS-95采用DS-CDMA及相移键控调制[12-13]。另外，蜂窝频谱中的2GHz个人通信系统（personal communication system，PCS）的频段已经拍卖，购得频段的运营商可自行决定采用何种标准。这样最终的结果是这一频段上有3个不同的标准：IS-136、IS-95和欧洲的GSM标准。日本的数字蜂窝系统与IS-136相似，只是频率不同。欧洲的GSM和美国的GSM频率不相同。互不兼容的标准如此繁多，要想在国内或全球范围内实现漫游，只能用一部多模手机，或

1. 缩写词GSM原本表示法语中的“移动通信特别小组”（Groupe Spéciale Mobile），这是欧洲GSM标准起草机构的名称。随着GSM系统在全球的普及，人们将这个缩写词的含义改成了“全球移动通信系统”（Global Systems for Mobile Communications）。
2. IS-136是早期的IS-54的演进版，可归入IS-54。

者同时用多部手机（对应多个手机号码）。 12

第二代数字蜂窝系统后来都增强以支持高速分组数据业务[14]。GSM通过把所有时隙分配给一个用户的方法使数据速率达到140kbit/s，这种增强型版本称为GPRS。另一种更为彻底的GSM增强型是EDGE（Enhanced Data Services for GSM Evolution），它通过采用编码高阶调制，将数据速率提高到384kbit/s。因为高阶调制对衰落敏感，所以EDGE采用了自适应技术来抑制衰落的影响。EDGE具体定义了九种不同的调制编码组合，每一种适合于一定的接收信噪比范围。接收端测量出接收信噪比，将其反馈给发送端，发送端依据这个反馈选择适合的调制编码组合进行发送。IS-136系统也采用GPRS和EDGE技术使数据速率提高到384kbit/s。IS-95系统则是采用捆绑扩频码的方法使数据速率达到115kbit/s[15]。

第三代蜂窝系统（3G）都是基于宽带CDMA的，它由国际电信联盟（International Telecommunications Union，ITU）发起[14]，称为国际移动通信2000标准（International Mobile Telecommunications 2000，IMT-2000）。IMT-2000根据不同的移动性和用户位置提供不同的数据速率：步行384kbit/s、汽车144kbit/s、室内2Mbit/s。3G标准和2G的系统不能兼容，运营商必须要先投资建立一个全新的基础设施，然后才能提供3G业务。最早的3G系统出现在日本，一个原因就是日本的频谱分配是授予的，不需要太多的前期花费。而在欧洲和美国，3G频谱是由拍卖决定的，任何运营商要想提供3G业务，都要先花费高昂的初期投资获得频谱。欧洲公司在3G频谱的拍卖中总共花费了1000亿美元。人们对拍卖这种方式存在一些争议，不少公司抱怨竞拍的形式使买家出价远远高于频谱本身的价值，倘若不能从拍得的频谱中充分获利，公司就会陷入困境。少数公司已决定取消在3G频谱上投资，不再参与3G系统建设。欧洲3G系统发展的实际情况并不如预想的那么顺利。目前看来，起码在一段时间内，只要增强2G的数据能力就能充分满足用户的需要。不过欧洲的2G频谱过于拥挤，所以最终结果要么是用户逐步转到3G系统，要么就是允许将3G的频段给2G使用（目前这在欧洲是不允许的）。美国的3G发展远远落后于欧洲，美国的可用的3G频谱也只有欧洲的一半。因为对于3G该使用哪部分频谱还存在争议，所以美国的3G频谱拍卖迟迟没有进行。不过美国允许1G和2G的频谱用于3G系统。相比于欧洲来说，美国的这种灵活性可以使系统的过渡更为平稳。由此看来，美国3G频谱拍卖的迟缓倒可以让FCC及运营商吸取欧洲人和日本人的经验教训。 13

1.4.2 无绳电话系统

无绳电话自20世纪70年代后期出现以来发展迅速。如今许多美国家庭中只有无绳电话。这可能是一种安全隐患，因为和有线电话不一样，无绳电话停电时无法工作。无绳电话的原设计目的是提供一种低成本、低移动性的无线连接公众电话交换网（PSTN）方式，即用无线连接来替代电话机机座和手持电话这两部分之间的那个线绳。无绳电话的竞争对象是有线的手持电话机，所以它必须能提供与有线电话一样的话音质量。早期的无绳电话就因为话音质量很差而被用户抛弃。第一代无绳电话系统中，每个座机只允许连接一个手持机，覆盖区也仅限于家庭或办公室的几间房屋范围。美国今天还是主要

使用这种系统，虽然也有一些座机能支持多个手持机且覆盖范围有所扩大的系统。在欧洲和亚洲，数字无绳电话系统已经演进到可以覆盖很大的区域，不光能在家里用，离开家也能用，在许多方面和蜂窝系统非常相似。

无绳电话系统的机座如同有线电话一样以相同的方式连接到PSTN，因此电话网的建设并没有增加复杂性。无绳电话的移动范围非常有限，手持机只能在机座所覆盖的附近距离内移动。不同的无绳电话系统之间没有协调机制，因此在住宅公寓这类无绳电话密集的地方，有可能产生很大的系统间干扰。为此，现在的无绳电话系统设有多个信道，它能扫描各个信道，找出干扰最小的信道进行通话。许多无绳电话系统采用了扩频技术来抑制干扰，这些干扰可能来自其他无绳电话系统，也可能来自婴儿监视器、无线局域网等其他系统。

在欧洲和亚洲，第二代数字无绳电话系统（cordless telephone, second generation, CT-2）把覆盖范围扩展到了室外。CT-2在室内使用时和传统无绳电话没什么区别。在室外，CT-2把基站（又称电话点或电信点）安装到人口密集区，如购物商场、繁华街道、火车站、飞机场等。只要无绳电话机是某个电信点提供商的用户，它就可以在该电信点的覆盖区内任意拨打电话。但不能从电信点呼入到无绳电话，因为网络不具有对移动用户的路由支持。为了弥补这个不足，有些CT-2电话内置了一个寻呼机。CT-2的用户在不同电信点覆盖区内移动时，呼叫不会切换，因此用户一旦拨打电话，就必须要待在这个电信点的覆盖区内，直至通话结束。CT-2业务曾经两次引入英国，但两次都以失败告终。不过20世纪90年代中期在中国香港特别行政区和新加坡取得了迅速的发展，但几年后因为蜂窝运营商通过降低资费加入竞争，使CT-2的增长势头迅速变弱。人们对CT-2的主要不满是覆盖不完善及不能切换。而蜂窝网络不存在这种问题，所以当蜂窝网的资费也很有竞争力时，人们就不会选择CT-2。目前CT-2业务大部分已经消失。

14

无绳电话的另一个演进是欧洲的DECT（Digital Enhanced Cordless Telecommunications）系统，它主要是针对办公楼设计的。DECT的主要功能是为一幢大楼内的用户交换机（private branch exchange，PBX）的电话使用者提供局部的移动性。DECT系统在楼内安装了许多基站，这些基站通过控制器与PBX相连。手持机和距离最近的基站通信，当用户走到另一个基站的覆盖区域时，系统能进行呼叫切换。DECT系统的用户也可以接听电话。DECT标准也支持CT-2的电信点业务，也许是因为CT-2业务的失败，这项功能并没有引起太多关注。欧洲目前大约有700万DECT用户，这个标准目前还没有扩展到其他国家。

更为先进的无绳电话系统是日本的PHS（Personal Handyphone System）[1]。PHS和蜂窝系统非常相似，拥有广泛分布的基站，能支持基站间的切换和呼叫路由。这使得PHS的发展要比CT-2顺利得多。作为一种新技术，PHS经历了前所未有的快速增长。在其出现两年之后的1997年，PHS的用户数达700万之多，但之后开始下降，原因是蜂窝运营商大幅度降价。尽管受到了蜂窝系统的冲击，PHS的平价资费和相对较高的数据速率（128kbit/s）仍然吸引了不少用户，到2005年PHS还有400万用户。PHS运营商正准备把

1、即小灵通。——译者注

数据速率提高到1Mbit/s，这样，目前的蜂窝系统还无法与之匹敌。PHS和蜂窝系统的主要区别在于，以车速运动时，PHS不能支持切换，这个缺点主要是PHS所用的动态信道分配技术引起的。动态信道分配能大大增加单基站可支持的用户数量及数据速率，因此可显著降低系统成本，但同时也使切换的过程变得复杂。假如PHS能保持其受欢迎的程度，那么它不大可能重蹈CT-2的覆辙，特别在其数据速率得到提升以后，更是如此。

从无绳电话系统这些年来的发展情况可以看到，让无绳电话从室内走到室外所需要的，要么是具有和蜂窝系统相类似或更先进的功能，要么就是低廉的成本。

1.4.3 无线局域网

当用户在一个较小区域，如校园或不大的建筑物内移动时，无线局域网能提供高速数据传输。接入无线局域网的无线设备通常都是静止的或是慢速移动的。在美国，所有无线局域网标准都工作在开放频段上，主要包括900MHz、2.4GHz、5.8GHz的ISM频段以及5GHz的U-NII（Unlicensed National Information Infrastructure）频段。在ISM频段上，所有开放业务都是次要业务。所以当主要业务运行时，开放业务会受到干扰。U-NII频段的好处就是不存在主要业务。ISM和U-NII频段虽然有无需FCC许可的优点，但这个优点也是一把双刃剑，因为其他开放系统也同样可以工作在这些频段上，这就会在系统间造成非常大的干扰，因此需要限制开放系统的功率谱密度来减轻干扰。无线局域网既可以采用星型结构，也可以采用对等网络结构（P2P, peer to peer）。在星型网络结构中，无线终端连接到覆盖区中的无线接入点或集线器，P2P结构中无线终端以自组织方式组网。 [15]

20世纪90年代初，由于人们对无线高速数据业务有着“压抑已久的渴望”，无线局域网公司和产品不断涌现。第一代无线局域网是专网性质的，协议互不兼容，大多工作在900MHz ISM频段内的26MHz带宽上，采用直序扩频技术，数据速率大约是1Mbit/s～2Mbit/s，既有星型结构又有P2P结构。缺乏标准化造成高昂的开发成本、低产量以及很小的市场份额，使得这些早期产品中只有少数获得了一定的成功。第一代无线局域网中，工作在900MHz频段之外的只有摩托罗拉的Altair，它的工作频率是18GHz的授权频段。Altair的数据速率大约是6Mbit/s。18GHz上的器件成本及路径损耗都很高，这使Altair的性能大受影响，它在问世几年之后就消失了。

美国的第二代无线局域网工作在2.4GHz的ISM频段上，占用83.5MHz的带宽。为了避免第一代无线局域网中出现过的不兼容问题，IEEE专门制定了相应的无线局域网标准，即IEEE 802.11b标准。802.11b采用直序扩频技术，数据速率大约是1.6Mbit/s（11Mbit/s的原数据速率），覆盖范围约为100m。网络多数是星型结构，也有少数是P2P结构。很多公司开发了基于802.11b的无线局域网产品。经过初期的缓慢增长后，基于802.11b的产品现在已经比较流行。许多笔记本电脑都已集成了802.11b无线网卡。许多公司和大学都建立了802.11b基站，许多咖啡厅、机场、酒店提供免费的无线局域网接入。

为了提供比802.11b更高的数据速率，802.11系列中又新开发了两个标准：802.11a和802.11g。802.11a工作在5GHz的U-NII频段上，占用带宽300MHz，它通过多载波调

制技术能在30m的覆盖范围内提供54Mbit/s的数据速率。802.11a占用更宽的带宽，所以拥有更多的信道数，可支持更多的高速用户。起初人们还担心802.11a会比802.11b贵很多，但很快就发现802.11a的价格也相当有竞争力。802.11g的设计与数据速率和802.11a相同，只是工作频段在2.4GHz，覆盖范围是50m。出于兼容性的考虑，许多无线网卡和接入点同时支持这3种标准。

无线局域网在欧洲的发展一直围绕着HIPERLAN（high-performance radio LAN）进行。HIPERLAN/2标准和IEEE 802.11a无线局域网标准很相似。两者的链路层设计很相像，HIPERLAN/2也工作在5GHz的频段上，类似于U-NII频段。它的最大数据速率及覆盖范围也和802.11a一样，分别是54Mbit/s及30m。它和802.11a在接入层协议及服务质量（Quality-of-Service，QoS）支持上有所不同。

1.4.4　广域无线数据业务

广域无线数据业务为高移动性的用户在广大的覆盖范围内提供无线数据传输。这些系统通过一个架设在高塔、屋顶或山顶之上的基站为特定的地理区域提供服务。这些基站既可以和有线骨干网相连，也可以形成一个多跳的无线自组织网络。

16

最初的广域无线数据业务提供的数据速率非常低，还不到10kbit/s，之后逐渐增加到20kbit/s。这项业务有两个主要提供商：Motient和Bell South Mobile Data（前身是RAM Mobile Data）。Metricom也提供相似的业务，但它是用大量低价格小型基站覆盖整个服务区，每个基站只覆盖一个较小的范围。减小基站覆盖区提高了频谱效率，这使Metricom提供的数据速率可达到76kbit/s，大大超过了其他广域无线数据系统。然而，高昂的基础设施成本最终导致Metricom公司破产，系统也随之停止运行。部分基础设施被收购后以品牌Ricochet在少数地区运行。

CDPD（cellular digital packet data）系统是叠加在模拟蜂窝网上的一种广域无线数据业务。由于数字蜂窝系统快速增长，许多模拟蜂窝系统的话音信道处于空闲状态，因此CDPD可以共享这些FDMA话音信道。CDPD业务能在全美范围内提供19.2kbit/s的分组数据传输。不过由于新一代蜂窝系统也能提供数据业务，传输速率甚至更高，因此CDPD几乎被完全替代。

除上述系统外，还新出现了一些能提供宽带接入的系统。总地看来，广域无线数据业务并不是很成功。

1.4.5　宽带无线接入

宽带无线接入可以在一个固定的接入点和若干终端之间提供高速无线通信。此类系统起初的设计目的是用来支持交互式家庭视频业务，后来业务的重点很快转移到高速（数十兆比特每秒）互联网接入及其他家庭或企业的高速数据网。在美国，此类系统可以在两个频段上使用：部分28GHz频谱留给本地多点分配业务（local multipoint distribution service，LMDS）；2GHz频谱留给多通道多点分配业务（multichannel multipoint distribution services，MMDS）。LMDS使新的运营商快速进入竞争激烈的宽

带服务领域，详见文献[5]的2.3节。MMDS可为广播电视和长途电话提供服务，它的有效传输距离是30km～50km，参见文献[5]的11.11节。MMDS能够承载上百路的数字电视节目、电话及互联网接入业务，主要竞争对手是现有的有线电视和卫星系统。欧洲也正在开发一种和MMDS相类似的Hiperaccess标准。

WiMax是基于IEEE 802.16的一种新的宽带无线技术[16~17]。802.16是一个宽带无线接入标准，工作在2GHz～11GHz的非视距传播条件，以及10GHz～66GHz视距传播条件。其数据速率对于固定用户约为40Mbit/s，对于移动用户约为15Mbit/s，覆盖范围可达几千米。许多笔记本电脑和个人数字助理（personal digital assistant，PDA）厂商已经准备在WiMax开通后在其产品中集成WiMax支持，以使其用户能够在任何地方随时接入互联网。WiMax的竞争对手是无线局域网、3G蜂窝系统，也可能包括有线业务，如有线电视和数字用户线路（digital subscriber line，DSL）。WiMax能否和这些系统竞争甚至替代它们将取决于它的性能和成本，这一切还是未知数。

17

1.4.6 寻呼系统

寻呼系统以非常高的功率（几百瓦甚至上千瓦）从多个基站或卫星上同时广播简短的寻呼消息。采用陆地发射机的系统通常可以覆盖一个特定的地理区域，例如一个城市；采用GEO同步卫星进行发射的系统可以覆盖一个国家乃至全球。寻呼消息是广播发送到整个覆盖区域的，所以不需要位置管理和路由功能。寻呼系统的发射机功率很高、复杂度也相对较高，从而使接收机可以做到复杂度低、耗电小和体积小巧，用一个轻便的电池就可以有很长的待机时间。此外高发送功率也使得寻呼信号很容易穿透建筑物的墙壁。寻呼业务不论是购机费还是月使用费，都要比蜂窝电话便宜，不过近年来这种价格优势在不断消失，因为蜂窝电话的价格在不断下降。寻呼系统的主要优势是低廉的价格、小而轻便的寻呼机、电池工作时间长、寻呼设备不受室内室外的限制。

早期的寻呼系统向用户发送1bit的消息，表示有人正在找他。用户通过有线电话回叫以获得对方的电话号码。后来发展到能发送简短的数字信息，包括向被呼方发送一个电话号码或者简短留言。无线寻呼在初期曾经大获成功，顶峰时期全美国有五千万用户。随着蜂窝电话的普及和其资费的下调，寻呼系统开始衰退。为了应付来自蜂窝电话的竞争压力，一些寻呼系统尝试增加新的功能。有些寻呼系统增加了“应答”功能（如双向通信），这就要求寻呼机的设计上的更改，因为它不只能接收消息，还必须能够发送消息，而且寻呼机到卫星或基站的距离又非常远。一些寻呼公司还尝试与掌上电脑厂商合作，让掌上电脑集成寻呼功能[18]。尽管有过这些努力，寻呼的市场还是在急剧缩水。只在医生及一些要求随时随地能找到人的行业中，寻呼还有一些市场。

1.4.7 卫星通信网

民用卫星通信是无线通信中的又一个重要分支[2~3]。采用GEO卫星的系统有Inmarsat和OmniTRACS。Inmarsat主要用于偏远地区的模拟话音传输，例如新闻记者可利用Inmarsat对战区进行实况报道。第一代Inmarsat-A系统的终端价格非常昂贵，体积也很

大，它采用1m口径的碟形抛物面天线。新一代Inmarsat系统采用数字技术降低了终端的成本和尺寸，大小和公文包差不多。高通公司的OmniTRACS系统可提供定位功能及双向通信功能，主要用于传送文本消息以及车队的位置跟踪。通过GEO卫星传送话音和数据有很多困难。信号要想到达卫星就需要有很大的发射功率，这样手机免不了都是又大又笨的。另一个问题就是在双向话音通信中，往返时延较大，这样大的时延能被用户明显感觉到。GEO卫星的数据速率一般低于10kbit/s。这些原因使人们认为轨道较低的LEO卫星更为适合语音及数据通信。

18

LEO系统需要30颗～80颗卫星才能实现全球覆盖。20世纪90年代后期，部署此类系统的计划层出不穷，其中最为雄心勃勃的铱星系统就是在此期间发射的。LEO系统建造、发射、维护这些卫星的花费要远远超过蜂窝系统中地面基站的费用。LEO系统在人口稀少的地区可以作为陆地通信系统的补充，还能做到只用一部手机、一个号码就能全球漫游，但是蜂窝系统的快速发展和资费的降低阻碍了LEO系统的发展壮大。铱星系统最终走向破产，其他许多计划中的系统根本就没有启动。只有全球星系统（Globalstar）是个例外，它目前在广大的覆盖范围内提供数据速率低于10kbit/s的话音和数据服务。另外，铱星系统的部分卫星也还在工作。

卫星系统最具吸引力的用途就是在大范围内广播视频和音频节目。在美国，每8个家庭中就有1个拥有卫星直播业务，卫星收音机也开始普及。类似的音频和视频卫星广播业务在欧洲也很普遍。卫星是广播的最好手段，因为它可以覆盖很大的区域，传播时延也不是问题。若考虑建设资金可以分期至多年，并可由众多用户分担，那么卫星系统在成本上和地面广播娱乐系统相比将非常有竞争力。

1.4.8 低成本、低功率的无线通信：蓝牙和紫蜂

随着成本和功率的不断降低，无线通信可以被嵌入到多种电子设备中，用来实现智能家居、传感器网络及其他应用。在这种趋势下出现了两种无线通信方式：蓝牙和紫蜂。

蓝牙[1]（Bluetooth）依靠简易的网络功能为无线电子设备提供短距离无线连接。蓝牙标准把一块集成了无线收发信机的很小的微芯片植入数字设备中，如手机、笔记本电脑、掌上电脑、打印机、投影仪、网络接入点等，用收发信机替代了这些数字设备原本需要的连线，例如笔记本电脑和打印机之间的连线、手机和无线耳机之间的连线。以1mW功率发射时，蓝牙的通信距离是10m，若将功率增大到100mW，通信距离可延伸至100m。蓝牙系统工作在2.4GHz开放频段上，可在世界范围内使用而无需考虑许可问题。蓝牙提供一个速率为723.2kbit/s的异步数据信道以及一个数据速率为57.6kbit/s的反向信道，此模式称为异步无连接（Asynchronous Connection-Less，ACL）。蓝牙也允许同时有3个速率为64kbit/s的同步信道，此模式称为同步面向连接（Synchronous Connection Oriented，SCO），主要用于支持耳机之类的话音传输，也可用于数据传输。

1. 蓝牙标准得名于公元940年至985年在位的丹麦国王哈拉德·布鲁图斯一世（Harald I Bluetooth），他统一了丹麦和挪威。蓝牙志在统一设备之间的无线连接接口，故取此名。

这些不同模式的总比特速率大约是1Mbit/s。异步数据经由分组交换协议传输，该协议采用每秒1600跳的调频技术。同步数据采用电路交换协议传送。

19

蓝牙采用频率间隔为1MHz的跳频多址接入技术，80MHz的总带宽分为80个不同的频率，任一时刻的传输带宽都是1MHz，这个传输带宽可由8个蓝牙设备同时使用，不同设备对应不同的逻辑信道，也即不同的跳频序列，它们共享这80MHz的带宽。处于不同微微网、不同逻辑信道上的蓝牙设备可能因为使用相同的跳频频率而发生频率碰撞。碰撞次数随微微网数目增加，系统性能也相应下降。

蓝牙标准由3Com、爱立信、英特尔、IBM、朗讯、微软、摩托罗拉、诺基亚和东芝等公司共同发起。目前已有超过1300家制造商采用该标准，已经推出了许多装备有蓝牙的消费类电子产品，包括手机的无线耳机、无线USB或RS232连接器、无线PCMCIA卡和无线机顶盒等。

紫蜂[1,2]（ZigBee）的设计目标是要比蓝牙有更低的成本和更低的功耗[19]。紫蜂基于IEEE 802.15.4标准，它和蓝牙一样工作在ISM频段。每个紫蜂网络可容纳255个设备，最大覆盖范围30m，最大数据速率是250kbit/s。虽然紫蜂的数据速率比蓝牙低，但它发射功率大大低于蓝牙，同时覆盖范围也更大。紫蜂希望能做到几个月甚至几年不用充电。紫蜂的应用目标包括传感器网络、存货标签等。

1.4.9 超宽带无线通信

超宽带（Ultrawideband，UWB）无线通信的频谱非常宽，因而在高速数据传输方面很有潜力[20-21]。超宽带通信的概念源于马可尼的火花隙发射机，其信号占用了很大的带宽。让单个低速率用户占用很大频段自然是很低效的，所以这种宽带通信很快被效率更高的技术所取代。直到2002年人们对它的兴趣又重新燃起，这是因为FCC允许UWB系统可以和原有用户重叠使用3.1GHz～10.6GHz的频段。从理论上来说，重叠使用将会干扰这一频段上的所有系统，其中包括安全和军事部门的系统、使用开放频率的802.11无线局域网和蓝牙等系统，以及花巨资买到频率专用权的蜂窝系统。因为损害到原有用户的利益，FCC的决策引起了很大的争议。为了尽量减小UWB系统对原有用户的影响，FCC对UWB的发射功率做出了严格的限制，使得UWB的收发设备只有靠得很近才能通信。

UWB有一个独特之处是雷达和通信界非常喜欢的，这就是其宽带特性拥有精确的测距能力。UWB的宽频带使它在超高速数据传输方面很有潜力。UWB系统的功率限制也使其设备的功耗和体积都能做得很小。

20

最初的UWB系统采用超窄脉冲进行简单的脉幅或脉位调制。多径效应会严重影响此类系统的性能但其影响可运用均衡及多载波调制来减弱。精确而快速的同步是UWB技术所面临的一大挑战。尽管还有许多问题有待解决，UWB在商业界和学术界都引起了人们巨大的兴趣。

1. 紫蜂得名于蜜蜂之舞。蜜蜂在发现新的食物后以飞舞的方式告诉其同伴。
2. 紫蜂的紫是ZigBee中Zig的音译，ZigBee中的Zig来自于蜜蜂飞舞的曲线呈之字形。——译者注

1.5 无线频谱

1.5.1 频谱的分配方式

绝大多数国家都设有专门的政府机构来负责分配和控制无线频谱的使用。在美国，商用频谱由联邦通信委员会（FCC）负责，军用频谱由频谱管理局（Office of Spectral Management，OSM）负责。在欧洲，商用频谱分配由欧洲电信标准化协会（European Telecommunications Standards Institute，ETSI）负责。国际的频谱分配由国际电信联盟（International Telecommunications Union，ITU）负责。政府决定分配多少频谱给商用和军用，这种分配会依据需要而有所变化。FCC过去的做法是先给各频谱指定用途，再向使用这一频谱的公司或团体发放许可证。例如FCC在20世纪80年代将800MHz频段指定给模拟蜂窝电话业务，然后按照一定条件在每一地区向两家运营商发放许可证。今天，FCC以及其他国家的管理机构仍旧给不同的频段指定具体的用途，但在频率的使用者方面，现在普遍拍卖给出价最高者。对于政府该怎样分配有限的频谱资源这个问题，社会上有一些争议。有些人认为，基于市场的拍卖方法最为公平也最有效率，还能给政府带来巨额收入，另一些人则认为，拍卖的方式将扼杀创新、限制竞争、不利于技术改进。拍卖使得只有大公司或大集团才能买得起频谱，而公司在购买频谱上花费了大量资金，势必影响系统的迅速建设开通，这将最终导致终端用户的成本升高。欧洲3G频谱拍卖中，出现了一些公司最终无力支付频谱费用的情形，这也助长了反对频谱拍卖的声音。

除拍卖外，一些特定频段被作为开放频段而留出，只要符合一定的行业规定，就可以无需许可而免费使用。相关的行业规定一般包括特定的标准、功率电平等。设置这些*开放频段*（unlicensed band）的目的是鼓励创新和降低成本。许多非常成功的无线系统就是工作在开放频段的，包括无线局域网、蓝牙和无绳电话。开放频段的最大问题恰恰就是它会因为自己的成功而被毁掉：如果一些业务很成功，那么在同一开放频段上就会有大量设备，如果这些设备距离比较靠近的话，它们就会相互干扰，大量干扰将使这个频段无法使用。

另一种分配频谱的方式是重叠，就是在已经分配了的频谱上重复分配一个业务作为次要业务，原有的业务称主要业务。一般会对次要业务的功率谱密度等做出严格限制，以尽量减小它对主要业务的影响。UWB就是重叠在ISM开放频段上的。具体确定次要业务对主要业务的干扰程度是一个复杂的问题，这也使重叠分配的方式引起了很大的争议。不过出于频谱资源稀缺的无奈，为了给新的系统和新的业务分配频率，重叠的做法已越来越普遍。

卫星系统覆盖的区域可以横跨多国乃至全球。对于这种横跨多国的无线系统，频谱分配是由国际电信联盟无线电部门（International Telecommunications Union Radio Communications group，ITU-R）负责的。该组织的另一部门ITU-T负责制定全球系统的电信标准，以保证不同国家的系统可以互通。

自20世纪初有频谱管理机构以来，频谱分配的模式基本上没有什么变化，新出现的

拍卖、重叠也没有从本质上改变频谱分配的模式，目前，全世界的频率管理机构中正在兴起一个试图改变频谱分配方法的浪潮。其目的是依靠先进的无线技术成果，使频率分配方式更为灵活和高效。颇具吸引力的一个设想就是智能无线电或称感知无线电，它通过感知周围的频谱环境，找出一个时间、空间、频率的范围，在此范围内即使以中高功率发射也不会对其他用户产生干扰。把这种方式用在很宽的频带上，就能产生出大量新的可用带宽，为新的无线系统和应用提供大量机会。不过，实现这种频谱分配上的彻底的改变还需要克服许多技术和政策上的障碍。

1.5.2 现有系统的频谱分配

绝大多数无线通信系统工作在30MHz～40GHz的频率范围内。这段频谱是无线通信较为理想的工作频段，它不受地球表面曲率的影响，天线大小适度，还可以穿透电离层。值得注意的是，良好接收所需要的天线尺寸与信号频率成反比，因此提高频率就能使天线更小。但用全向天线接收时，接收信号功率和频率的平方成反比，因此高频信号很难覆盖大距离通信。

如前所述，频谱分配可能是授权式（由管理机构分配给特定的运营者），也可能是开放式（只要满足一定的要求，谁都可以用）。表1-1列出了美国目前主要商用无线系统的授权频谱分配，欧洲和亚洲的频谱分配也与此类似。

表1-1 美国的授权频谱分配

业务/系统	频率范围
AM广播	535kHz~1605kHz
FM广播	88MHz~108MHz
电视广播（2~6频道）	54MHz~88MHz
电视广播（7~13频道）	174MHz~216MHz
电视广播（UHF）	470MHz~806MHz
无线宽带	746MHz~764MHz，776MHz~794MHz
3G无线宽带	1.7GHz~1.85GHz，2.5GHz~2.69GHz
1G、2G蜂窝系统	806MHz~902MHz
个人通信系统（2G蜂窝电话）	1.85GHz~1.99GHz
无线通信业务	2.305GHz~2.32GHz，2.345GHz~2.36GHz
数字卫星通信	2.32GHz~2.325GHz
多通道多点分配业务（MMDS）	2.15GHz~2.68GHz
数字电视直播卫星	12.2GHz~12.7GHz
本地多点分配业务（LMDS）	27.5GHz~29.5GHz，31GHz~31.3GHz
无线固定业务	38.6GHz~40GHz

需要注意的是，数字电视和广播电视所用的频段是一样的，因此所有的电视广播设备最终都要完成模拟到数字的转变。另外，3G无线宽带的频谱目前被划分到UHF电视60~69频道，即将重新分配。1G模拟和2G数字蜂窝系统都工作在800MHz的频段上，由运营商决定数字和模拟的频率各占多少。

开放频段由各国管理机构负责分配。各国通常对开放频谱的分配使用是相互协调的，22 这样有利于技术规范在世界范围内兼容。表1-2列出了美国开放频谱的分配情况。

表1-2 美国的开放频谱

频 段	频率范围
ISM I波段（无绳电话、1G无线局域网）	902MHz~928MHz
ISM II波段（蓝牙、802.11b及802.11g）	2.4GHz~2.4835GHz
ISM III波段（无线用户交换机）	5.725GHz~5.85GHz
U-NII I波段（室内系统、802.11a）	5.15GHz~5.25GHz
U-NII II波段（短距离室外系统、802.11a）	5.25GHz~5.35GHz
U-NII III波段（长距离室外系统、802.11a）	5.725GHz~5.825GHz

在ISM波段I上，授权用户的发送功率较高，同开放用户之间形成了相互干扰。因此这一频段对开放用户的限制比较严格，其性能也较差。整个U-NII波段由三个单独的100MHz波段构成，总带宽是300MHz，这3个波段上的功率限制有少许不同。有许多开放系统工作在这些频段上。

1.6 标准

通信系统之间要想互连互通，就需要标准化。标准通常是由全国性或国际性的组织制定的。美国的通信工业协会（Telecommunications Industry Association，TIA）就是这样的一个组织。这些组织最终采纳由其他机构提出的标准。在美国，IEEE是提出标准的主要机构，类似于欧洲的ETSI。这些机构通过一个漫长的过程来形成标准，这个过程需要厂商和其他相关单位的投人，需要经过反复的讨论。虽然形成标准需要花费大量时间，但为了能在开发系统时占到先机，厂商都积极参加，以期望他们的想法能被纳入标准。标准化的主要目的只是为了系统的互通性和互操作性，所以标准一般不会包括系统设计的所有细节，这也是为了能促进厂商进行创新，并能使一个厂商的产品可以区别23 于其他厂商的产品。

除了保证兼容性，标准的益处还在于能形成规模经济从而降低成本。例如，无线局域网工作在开放频段上，它们并不需要一定遵循某个特定的标准。第一代无线局域网没有统一的标准，许多系统都需要各自的专用部件，结果就是系统的成本居高不下，再加上性能也不佳，造成没有多少人愿意使用。以后的无线局域网汲取了这个教训，大力推进标准化进程，形成了非常成功的IEEE 802.11系列协议。

标准化当然也有缺点。首先，标准的制定过程并不完善，参与的公司都有各自的算盘，形成的结果往往既不能体现最好的技术，也不能给消费者带来最大的实惠。此外，标准化过程必然要在某个时候完成，此后要想再加入某些创新或改进都是非常困难的。最后，标准的出台往往带有政治色彩。例如美国的第二代蜂窝电话最终采纳了两个不同的标准，多少有点儿怪异。正是这一点导致美国在2G蜂窝系统的发展上大大落后于欧洲。虽然有这些缺陷，标准化对于无线系统的设计和运行不仅十分必要，也是非常有利

的。倘若所存在的这些问题可以被克服，那么就能给无线通信产业带来更大的好处。

习题

1.1 存储容量的不断增加使我们可以在体积更小的存储器内存入更多的数据。完全可以设想将大量的数据存入一个微芯片内。假设我们需要把这些数据传送到远处，一种方法是用卡车把存储数据的芯片拉到目的地，另一种方法是通过电通信的方式发送到目的地。请比较这两种方法的优缺点。

1.2 请说出至少2个无线系统中突发数据传输相对于连续数据传输的优点和缺点。

1.3 光缆传输的误比特率大约是10^{-12}。在某些无线信道中，DPSK调制的误比特率是$1/(2\bar{\gamma})$（$\bar{\gamma}$为平均信噪比）。要使无线信道的误比特率和光缆一样，请计算所需的平均信噪比。这样算得的信噪比要求非常高，因此无线信道的误比特率通常远大于10^{-12}。

1.4 分别就LEO、MEO、GEO的情形计算卫星与地球之间发送数据时的往返时延，假设光速为3×10^8m/s。若某话音系统可接受的最大往返时延是30ms，那么这几种卫星系统中哪一种适合于双向话音通信？

24

1.5 哪些应用可以显著增加对无线数据的需求？

1.6 本题展示一些运营商从单一话音业务转向混合业务时会遇到的经济问题。设有一个运营商拥有120kHz的带宽，他可以把这些带宽分配给话音用户或数据用户。每个话音用户需要20kHz的带宽，每个数据用户需要60kHz的带宽。如果把频谱全都分给话音用户可以得到6个话音信道，或者可以把所有带宽分成1个数据信道和3个话音信道，或者是其他分法。再假设这是一个时分的系统，每个时隙持续T秒。所有的话音和数据请求只在时隙的开始时刻发生，它们需要的通信时间都是T秒。假设系统中有6个独立的话音用户，每个用户以0.8的概率请求话音信道，每次通话费是0.2美元。系统中还有两个数据用户，每个用户以0.5的概率请求数据信道，每次通信费用是1美元。请分析应该如何分配带宽才能获得最大收益？

1.7 请举出使用无线局域网相对于使用有线局域网的3种缺点。对于什么样的应用，无线移动性这个优点会大于这些缺点？对于什么样的应用，缺点要大于优点？

1.8 蜂窝系统正在改用更小的小区来增大系统容量，这势必会导致设计的复杂化，列举至少3个这样的设计问题。

1.9 为什么最小化复用距离可以最大化蜂窝系统的频谱效率？

1.10 本题说明小区缩小对系统容量增加的影响。考虑一个面积为100km²的正方形城市，要求为这个城市设计一个采用正方形小区的蜂窝系统。每个小区不论大小都有100个信道，因而可支持100个用户同时通话（只要传播模型和发送功率适当，每小区可支持的用户数基本上与小区大小无关）。

(a) 如果小区面积是1km²，系统能支持多少个用户同时通话？

(b) 如果要求系统能支持250 000个用户同时通话，小区面积应该是多少？

注意用户并不是永远都在通话，现在我们来考虑一下这一点的经济含义。假设周五的下午5点到6点是蜂窝电话用户最繁忙的时段。在此期间，平均每个用户都要拨打1个电话，每个通话持续2min。假设系统要求用户在此繁忙时段的通话阻塞率不大于2%（阻塞率可用爱尔兰B模型计算：$P_b=(A^C/C!)/(\sum_{k=0}^{C}A^k/k!)$，其中$C$是信道数，$A=U\mu H$，$U$是用户数，$\mu$是

单位时间内平均发生的呼叫次数，H是平均每个呼叫的持续时间（见文献[5]的3.6节）。

（c）宏小区（1 km^2的小区）系统可以支持的用户数是多少？采用微小区，即面积为（b）小题结果的小区时，系统的用户数又是多少？

（d）如果一个基站的造价是50万美元，这两个系统（宏小区/微小区）基站部分的总造

25

价分别是多少？

（e）如果用户的月使用费是50美元，这两个系统的月收入分别是多少？它们分别需要多少时间才能收回基站部分的建设成本？

1.11　多少条CDPD数据线的数据速率才能相当于WiMax的平均速率？

参考文献

[1] V. H. McDonald, "The cellular concept," *Bell System Tech. J.*, pp. 15–49, January 1979.
[2] F. Abrishamkar and Z. Siveski, "PCS global mobile satellites," *IEEE Commun. Mag.*, pp. 132–6, September 1996.
[3] R. Ananasso and F. D. Priscoli, "The role of satellites in personal communication services," *IEEE J. Sel. Areas Commun.*, pp. 180–96, February 1995.
[4] D. C. Cox, "Wireless personal communications: What is it?" *IEEE Pers. Commun. Mag.*, pp. 20–35, April 1995.
[5] T. S. Rappaport, *Wireless Communications – Principles and Practice*, 2nd ed., Prentice-Hall, Englewood Cliffs, NJ, 2001.
[6] W. Stallings, *Wireless Communications and Networks*, 2nd ed., Prentice-Hall, Englewood Cliffs, NJ, 2005.
[7] K. Pahlavan and P. Krishnamurthy, *Principles of Wireless Networks: A Unified Approach*, Prentice-Hall, Englewood Cliffs, NJ, 2002.
[8] A. J. Goldsmith and L. J. Greenstein, "A measurement-based model for predicting coverage areas of urban microcells," *IEEE J. Sel. Areas Commun.*, pp. 1013–23, September 1993.
[9] K. S. Gilhousen, I. M. Jacobs, R. Padovani, A. J. Viterbi, L. A. Weaver, Jr., and C. E. Wheatley III, "On the capacity of a cellular CDMA system," *IEEE Trans. Veh. Tech.*, pp. 303–12, May 1991.
[10] K. Rath and J. Uddenfeldt, "Capacity of digital cellular TDMA systems," *IEEE Trans. Veh. Tech.*, pp. 323–32, May 1991.
[11] Q. Hardy, "Are claims hope or hype?" *Wall Street Journal*, p. A1, September 6, 1996.
[12] A. Mehrotra, *Cellular Radio: Analog and Digital Systems*, Artech House, Norwood, MA, 1994.
[13] J. E. Padgett, C. G. Gunther, and T. Hattori, "Overview of wireless personal communications," *IEEE Commun. Mag.*, pp. 28–41, January 1995.
[14] J. D. Vriendt, P. Lainé, C. Lerouge, and X. Xu, "Mobile network evolution: A revolution on the move," *IEEE Commun. Mag.*, pp. 104–11, April 2002.
[15] P. Bender, P. J. Black, M. S. Grob, R. Padovani, N. T. Sindhushayana, and A. J. Viterbi, "CDMA/HDR: A bandwidth efficient high speed wireless data service for nomadic users," *IEEE Commun. Mag.*, pp. 70–7, July 2000.
[16] S. J. Vaughan-Nichols, "Achieving wireless broadband with WiMax," *IEEE Computer*, pp. 10–13, June 2004.
[17] S. M. Cherry, "WiMax and Wi-Fi: Separate and Unequal," *IEEE Spectrum*, p. 16, March 2004.
[18] S. Schiesel, "Paging allies focus strategy on the Internet," *New York Times*, April 19, 1999.
[19] I. Poole, "What exactly is ... ZigBee?" *IEEE Commun. Eng.*, pp. 44–5, August/September 2004.
[20] L. Yang and G. B. Giannakis, "Ultra-wideband communications: An idea whose time has come," *IEEE Signal Proc. Mag.*, pp. 26–54, November 2004.
[21] D. Porcino and W. Hirt, "Ultra-wideband radio technology: Potential and challenges ahead," *IEEE Commun. Mag.*, pp. 66–74, July 2003.

26

第2章 路径损耗和阴影衰落

无线信道要成为一种可靠的高速通信媒质存在着严峻的挑战。无线信道易受噪声、干扰和其他信道因素的影响，而且由于用户的移动和信道的动态变化，这些因素还在随时间随机变化。本章讨论由路径损耗（path loss）和阴影（shadowing）效应所引起的接收信号功率随距离变化的规律。路径损耗是由发射功率的辐射扩散及信道的传播特性造成的。在路径损耗模型中一般认为对于相同的收发距离，路径损耗也相同。阴影效应是由发射机和接收机之间的障碍物造成的，这些障碍物通过吸收、反射、散射和绕射等方式衰减信号功率，严重时甚至会阻断信号。路径损耗引起长距离上（100m~1000m）接收功率的变化，而阴影引起障碍物尺度距离上（室外环境是10m~100m，室内更小）功率的变化。两者在相对较大的距离上引起功率变化，故称其为*大尺度传播效应*（large-scale propagation effect）。第3章所要讨论的多径信号干涉也会引起接收功率的变化，但这种变化发生在波长数量级距离上，这个距离较短，所以称为*小尺度传播效应*（small-scale propagation effects）。在路径损耗、阴影效应和多径传播的综合影响下，接收功率和发送功率比值的分贝值[1]（dB）与距离对数值的关系如图2-1所示。

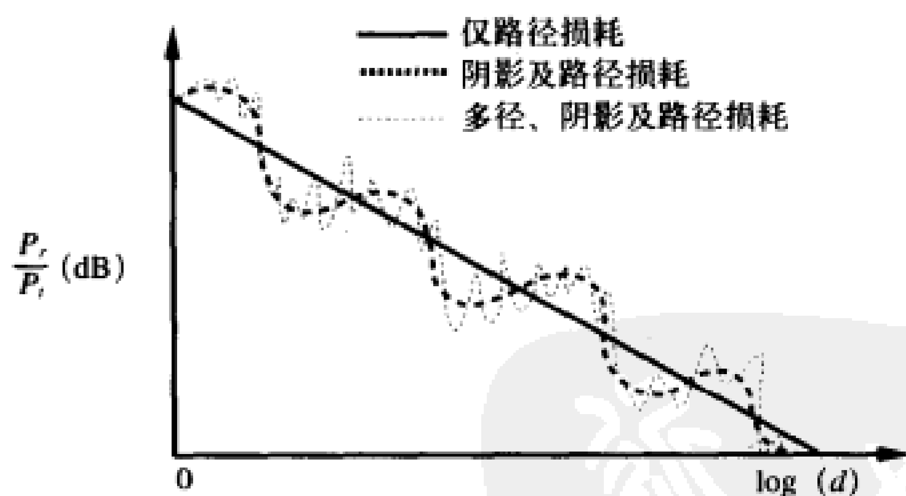


图2-1 路径损耗、阴影及多径与距离的关系

本章首先简要介绍电波传播，并给出信号模型，然后给出最简单的信号传播模型：自由空间路径损耗模型。两点之间的信号传播若无衰减和反射，则服从自由空间传播规律。接下来介绍射线跟踪传播模型，它可以根据麦克斯韦方程近似地描述电波传播特性，这种近似在多径分量很少且物理环境已知时是相当精确的。射线跟踪模型与信号传播区域内的地形和介电性质紧密相关。本章还将介绍一些经验模型，其参数是通过对室内和室外的信道进行测量后得到的。我们将给出一个简单的通用模型，它只有少量几个参数，但能很好地体现出路径损耗对系统分析的主要影响。本章还将给出基于大量障碍物的对数正态阴影衰落模型。如果多径分量很多或者传播环境的几何特性和介电性质未知，就

1. x的分贝值是$10\log_{10}x$。

必须用统计多径模型，我们将在第3章中对此进行介绍。

本章只是对路径损耗和阴影效应的信道模型进行了概述，而对各种频率下信道覆盖及传播模型问题的研究本身就能写一本书。这方面已经有一些优秀的教材，如文献[1~2]。多天线和超宽带等特定系统中的信道模型问题可参考文献[3~4]。

2.1　无线电波传播

人们对无线电波传播的最初了解要追溯到詹姆斯·克拉克·麦克斯韦（James Clerk Maxwell）的开创性研究。他于1864年建立的电磁波传播理论预言了无线电波的存在。1887年海因里希·赫兹（Heinrich Hertz）证实了电磁波的物理存在。然而他没有看到电磁波的实用特性，他认为由于声波频率非常低并且电磁波传播特性很差，因此无线电波不能用来携带语音。麦克斯韦和赫兹的研究开创了无线通信领域。1894年奥立弗·洛奇（Oliver Lodge）运用这些理论建造了第一个无线通信系统，尽管其传输距离只有150m。1897年企业家古列尔莫·马可尼（Guglielmo Marconi）成功地把无线电信号从英国的怀特岛发送到了18英里外的一艘拖船上。到1901年，马可尼的无线通信系统已经能够横跨大西洋。这些早期的无线通信系统使用电报信号进行通信。1906年，雷吉纳德·菲森登（Reginald Fessenden）使用幅度调制第一次进行了语音和音乐的传送，他把低频信号调制到高频电磁波上进行传输，从而突破了赫兹所说的低频传播限制，这也是今天的各种无线通信系统中普遍使用的方法。

28

电磁波在空中传播时，墙壁、地面、建筑物和其他物体会对电磁波形成反射、散射和绕射。电波传播的基本细节可通过求解带边界条件的麦克斯韦方程获得，其中边界条件表征了这些障碍物的物理特性。求解这样的麦克斯韦方程涉及大量复杂结构的雷达截面（radar cross-section，RCS）。相关的计算非常复杂，而且计算中必要的参数往往难以得到，因而人们采用了一些近似方法来描述信号的传播特性，避免求解复杂的麦克斯韦方程。

最常见的一种近似方法是射线跟踪技术。射线跟踪模型将波前近似为简单粒子，进而确定出反射和折射对波前的影响，但这种方法忽略了由麦克斯韦耦合微分方程组所描述的更为复杂的散射效应。最简单的射线跟踪模型是两径模型，指发射机和接收机之间只存在一条直射路径和一条反射路径的情况。两径模型能准确地反映信号的传播特性。两径模型中的反射信号一般是地面反射，对于高速公路及乡村道路和水面上的信号传播来说，两径模型是一种很好的近似。在接下来的几节中，我们将详细分析两径模型，以及增加了反射、散射和绕射分量的更复杂的模型。许多传播环境不能由射线跟踪模型准确描述，此时通常的做法是以实际测量结果为基础，建立经验模型。我们将讨论一些最常用的经验模型。

无线信道的复杂性和时变性往往使我们难以建立一个准确的确定信道模型，此时一般采用统计模型。如2.7节所述，信号传播路径上房屋等障碍物引起的信号衰减通常就是用统计的方法来表征的。统计模型也用来描述大量多径分量之间的相互干涉，如第3章所述。对于具有规则几何形状和均匀介电性质的传播环境，统计模型是最为精确的。

室内传播环境一般没有室外传播环境规则，因为不同的室内环境，如开放的工厂、办公室、金工车间等的几何形状和介电性质差别很大。对于上述传播环境可用计算机辅助建模工具来预测信号传播特性[5]。

2.2 发送和接收信号模型

我们主要研究频率范围为0.3GHz~3GHz的UHF频段和频率范围为3GHz~30GHz的SHF频段，这个频率范围的传播特性良好，天线尺寸也比较小，很适合于无线通信。我们假设传播距离不是很大，这样地球曲率的影响可以忽略。

调制器中的振荡器产生的是实正弦信号，不是复指数信号。所以我们认为接收与发送的信号均是实信号，虽然为了简化分析，我们经常采用复数信道建模，但信道实际上只是改变了发送信号在不同频率处的幅度和相位，因此接收信号也是实信号。因此，我们所研究的发射和接收信号都是实信号。但为便于分析，常把实的发送和接收信号表示成一个复信号的实部，这就是带通信号的等效基带表示。本章也将采用这种表示。有关带通信号与系统的等效基带表示的内容详见附录A。

29

发送信号表示式为

$$\begin{aligned} s(t) &= \mathrm{Re}\{u(t)e^{j2\pi f_c t}\} \\ &= \mathrm{Re}\{u(t)\}\cos(2\pi f_c t) - \mathrm{Im}\{u(t)\}\sin(2\pi f_c t) \\ &= s_I(t)\cos(2\pi f_c t) - s_Q(t)\sin(2\pi f_c t) \end{aligned} \tag{2-1}$$

其中$u(t)=s_I(t)+js_Q(t)$是一个复基带信号，其同相分量为$s_I(t)=\mathrm{Re}\{u(t)\}$，正交分量为$s_Q(t)=\mathrm{Im}\{u(t)\}$，带宽为$B_u$，功率为$P_u$。$u(t)$称为$s(t)$的*复包络*（complex envelope）或*等效基带信号*（equivalent lowpass signal）。复包络得名于$u(t)$的振幅就是$s(t)$的振幅。$u(t)$的相位包括了载波相位移。对于带宽$B \ll f_c$的带通信号，等效基带表示使信号通过$u(t)$控制而与载波频率无关。发送信号$s(t)$的功率为$P_t=P_u/2$。

接收信号的表示与发送信号类似，只是叠加了噪声：

$$r(t)=\mathrm{Re}\{v(t)e^{j2\pi f_c t}\}+n(t) \tag{2-2}$$

其中$n(t)$为信道噪声。等效基带信号$v(t)$与$s(t)$所传输的信道有关。当$s(t)$经过时不变信道时，由附录A可知$v(t)=u(t)*c(t)$，其中$c(t)$是等效基带信道的冲激响应。时变信道的情形将在第3章中讨论。

式（2-2）中的接收信号包括两项，第一项是经过信道传播之后的发送信号，第二项是信道噪声。接收信号的信噪比（signal-to-noise power ratio，SNR）定义为第一项和第二项的功率比。由于信号的传播特性和噪声无关，为了突出所研究的信号传播问题，本章和第3章将忽略随机噪声分量$n(t)$。但噪声对后续章节中所研究的无线通信系统的容量和性能问题有至关重要的影响。

当发射机和接收机中有一方在移动时，接收信号将会有一个多谱勒频移$f_D=v\cos\theta/\lambda$，θ是入射波相对于移动方向的角度，v是接收机沿其移动方向移动的速度，$\lambda=c/f_c$是信号波长，$c=3\times10^8$m/s是光速。图2-2显示了与多谱勒频移相关的几何关系。在很短时间间隔Δt内，发射机或接收机的移动将产生$\Delta d=v\Delta t\cos\theta$的行程差，该行程差引起的相位

变化为$\Delta\phi = 2\pi v\Delta t\cos\theta/\lambda$，再由频率和相位的关系得到多谱勒频移为

30

$$f_D = \frac{1}{2\pi}\frac{\Delta\phi}{\Delta t} = \frac{v}{\lambda}\cos\theta \tag{2-3}$$

如果接收机朝着发射机的方向移动（$-\pi/2 \leqslant \theta \leqslant \pi/2$），多谱勒频移为正值，反之为负值。对于典型的车速（75km/h）和频率（约1GHz），多谱勒频移只有100Hz左右[6]，因此本章的自由空间和射线跟踪模型将忽略多谱勒频移，我们将在第3章统计衰落模型中考虑多谱勒频移的影响。

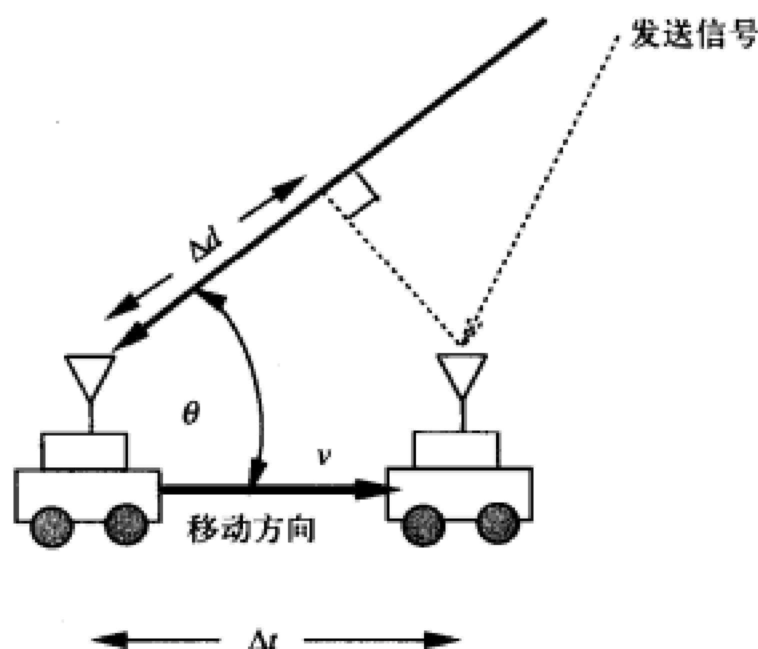


图2-2 多谱勒频移中的几何关系

假设发送信号$s(t)$的功率为P_t，相应的接收信号$r(t)$的功率为P_r。存在阴影时，P_r指对阴影平均后的结果。定义信道的**路径损耗真值**（linear path loss）为发射功率和接收功率的比值：

$$P_L = \frac{P_t}{P_r} \tag{2-4}$$

定义信道的**路径损耗**（path loss）为路径损耗真值的分贝数，即发送功率和接收功率的分贝差

$$P_L \mathrm{dB} = 10\log_{10}\frac{P_t}{P_r}\mathrm{dB} \tag{2-5}$$

信道是无源的，它只能衰减信号，所以用分贝表示的路径损耗一般都是非负值。定义**路径增益**（path gain）分贝值为路径损耗分贝值的负数：$P_G = -P_L = 10\log_{10}(P_r/P_t)\mathrm{dB}$，通常是负值。

考虑阴影效应时，物体的随机遮蔽将使接收功率呈现出随机性，2.7节将对此进行讨论。

2.3 自由空间路径损耗

假设信号经过自由空间到达距离d处的接收机，发射机和接收机之间没有任何障碍

物，信号沿直线传播。这样的信道称为视距（line-of-sight，LOS）信道，相应的接收信号称为LOS信号或直射信号。自由空间路径损耗使接收信号相对于发送信号引入了一个复数因子[1]，产生接收信号：

$$r(t)=\mathrm{Re}\left\{\frac{\lambda\sqrt{G_l}\mathrm{e}^{-\mathrm{j}2\pi d/\lambda}}{4\pi d}u(t)\mathrm{e}^{\mathrm{j}2\pi f_c t}\right\} \tag{2-6}$$

式中 $\sqrt{G_l}$ 是在视距方向上发射天线和接收天线的增益之积，$\mathrm{e}^{-\mathrm{j}2\pi d/\lambda}$ 是由传播距离d引起的相移。

31

发射信号$s(t)$的功率为P_t，由式（2-6）可得到接收功率和发射功率的比为

$$\frac{P_r}{P_t}=\left[\frac{\sqrt{G_l}\lambda}{4\pi d}\right]^2 \tag{2-7}$$

可见接收功率与收发天线间距离d的平方成反比。下节中将看到，在其他传播模型中接收功率随距离d下降的速度更快。式（2-7）中的接收功率还与波长的平方成正比，因此，载波频率越高则接收功率越小。接收功率与波长λ有关是因为接收天线的有效面积和波长有关[1]。若采用定向天线，接收功率也有可能随着频率的增加而增加[7]。

自由空间传播时，接收功率也可以表示为dBm的形式[1]

$$P_r\mathrm{dBm}=P_t\mathrm{dBm}+10\log_{10}(G_l)+20\log_{10}(\lambda)-20\log_{10}(4\pi)-20\log_{10}(d) \tag{2-8}$$

自由空间路径损耗（free-space path loss）定义为自由空间模型下的路径损耗：

$$P_L\mathrm{dB}=10\log_{10}\frac{P_t}{P_r}=-10\log_{10}\frac{G_l\lambda^2}{(4\pi d)^2} \tag{2-9}$$

相应的自由空间路径增益（free-space path gain）为

$$P_G=-P_L=10\log_{10}\frac{G_l\lambda^2}{(4\pi d)^2} \tag{2-10}$$

例2.1：有一室内无线局域网，载波频率$f_c=900\mathrm{MHz}$，小区半径10m，使用全向天线。在自由空间路径损耗模型下，如果要求小区内所有终端的最小接收功率为10μW，问接入点的发射功率应该是多大？如果工作频率变成5GHz，相应所需的发射功率又为多少？

解：应先求出终端位于小区边界时，达到最小接收功率所需要的发射功率。由式（2-7），计算所需发射功率的公式为

$$P_t=P_r\left[\frac{4\pi d}{\sqrt{G_l}\lambda}\right]^2$$

代入$G_l=1$（全向天线）、$\lambda=c/f_c=0.33\mathrm{m}$、$d=10\mathrm{m}$及$P_r=10\mu\mathrm{W}$，得$P_t=1.45\mathrm{W}=1.61\mathrm{dBW}$（注意dBW是相对于1W的分贝值，$P$W等于$10\log_{10}P$dBW）。5GHz时，$\lambda=0.06\mathrm{m}$，相应有$P_t=43.9\mathrm{kW}=16.42\mathrm{dBW}$。

32

1. dBm就是以W为单位的功率x表示成相对于mW的分贝值：$10\log_{10}(x/0.001)$。

2.4　射线跟踪

在典型的城市或室内环境中，从固定发射机发出的无线电信号在其传播环境中会遇到许多物体，这些物体将使发送信号发生反射、绕射和散射，如图2-3所示。经由反射、绕射和散射到达接收机的信号分量称为*多径信号分量*（multipath signal component）。接收端收到的多径信号相对于直射信号有功率衰减、时延、相移及频移。多径信号和直射信号在接收端叠加常常会造成信号失真。

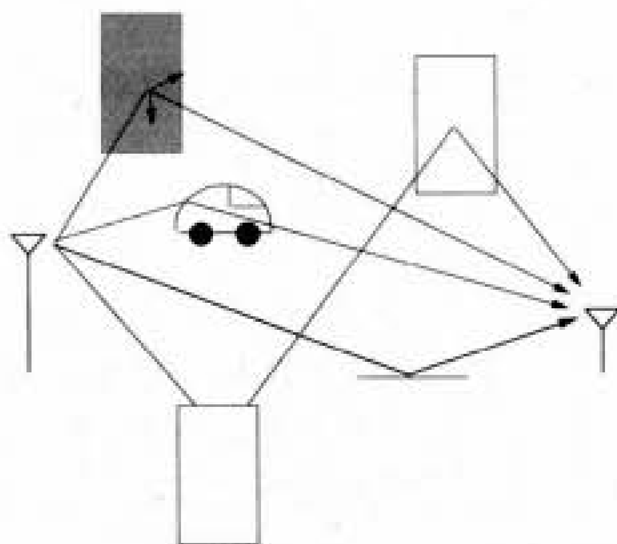


图2-3　反射、绕射与散射分量

在射线跟踪中，考虑有限个反射体，其位置和介电性质已知，那么在适当的边界条件下求解麦克斯韦方程就可以得到详细的多径传播过程。不过这种方法的计算太复杂，不适合成为通用的建模方法。射线跟踪法把波前近似为粒子，用一些简单的几何方程取代复杂的麦克斯韦方程，以此来近似反射、绕射和散射对波前的影响。当接收机和最近的反射体的距离有数个波长，所有反射体的大小相对于波长都足够大并且相当平滑时，射线跟踪法的近似误差最小。对比射线跟踪法的结果和经验数据可以发现，射线跟踪在乡村地区[8]、在发射机和接收机都很靠近地面的街道上[8-10]和在适当调整了绕射系数的室内环境中[11]，都能很精确地建模接收信号功率。而且除了接收功率变化外，射线跟踪法并不能有效地反映多径时延扩展等其他一些传播特性[12]。

33

如果发射机、接收机和反射体都是静止的，那么多个信号路径的特性就是固定的。如果发射机或接收机是运动的，这些路径的特性就是时变的。如果反射体的数量、位置和特性在任何时间都是已知的，就是确定的时变信道，否则就是随机时变信道，必须使用统计模型。类似地，如果反射体的数量很大或者反射体表面不光滑，也只能用统计模型来近似接收信号。我们将在第3章讨论统计衰落模型。文献中还有结合射线跟踪和统计衰落的混合模型[13-14]，这里将不进行讨论。

最一般性的射线跟踪模型包含直射、反射、绕射和散射等各种衰落分量，其建模需要知道发射机和接收机周围所有物体的几何和介电性质。相应的计算机软件有朗讯公司的WiSE（Wireless Systems Engineering）、Wireless Valley公司的SitePlanner®、马可尼公司的Planet® EV等。这些软件广泛用于室内和室外环境中的系统规则。对室外信道结合航摄

照片，对室内信道结合建筑制图，用这样的软件就能绘制出三维的传播环境几何图，见文献[5]。

以下几节将以从简到繁的次序讲述几种射线跟踪模型。首先讲述简单的两径模型，它反映地面反射波和直射波的干涉引起的信号变化。该模型适合于只有少量反射体的孤立区域，如道路，一般不适用于室内环境；接下来介绍适用于街道和走廊的十径反射模型；最后，我们给出一个可用来预测任意传播环境中信号传播的通用模型。两径模型只需要知道天线高度，十径模型需要知道天线的高度和街道/走廊的宽度，通用模型除需要这些参数外，还需要知道传播环境中反射体、绕射体和散射体的几何和介电性质。

2.4.1 两径模型

两径模型用在单一的地面反射波在多径效应中起主导作用的情形，图2-4所示是两径模型的示意图。接收信号由两部分组成：经自由空间到达接收端的直射分量和经过地面反射到达接收端的反射分量。

接收到的直射分量可由式（2-6）确定。反射路径由图2-4中的x和x'两段组成，若忽略反射面的衰减[1]，那么两径模型中的接收信号为

$$r_{2\text{-ray}}(t)=\mathrm{Re}\left(\frac{\lambda}{4\pi}\left[\frac{\sqrt{G_l}\,u(t)\mathrm{e}^{-\mathrm{j}2\pi l/\lambda}}{l}+\frac{R\sqrt{G_r}\,u(t-\tau)\mathrm{e}^{-\mathrm{j}2\pi(x+x')/\lambda}}{x+x'}\right]\mathrm{e}^{\mathrm{j}2\pi f_c t}\right) \tag{2-11}$$

其中$\tau=(x+x'-l)/c$是反射波相对于直射波的时延，$\sqrt{G_l}=\sqrt{G_aG_b}$是直射方向上发射和接收天线增益的乘积，R是地面反射系数，$\sqrt{G_r}=\sqrt{G_cG_d}$是x方向上的发送天线增益和x'方向上接收天线增益的乘积。两径模型的时延扩展（delay spread）等于反射路径相对于直射路径的时延：$(x+x'-l)/c$。

34

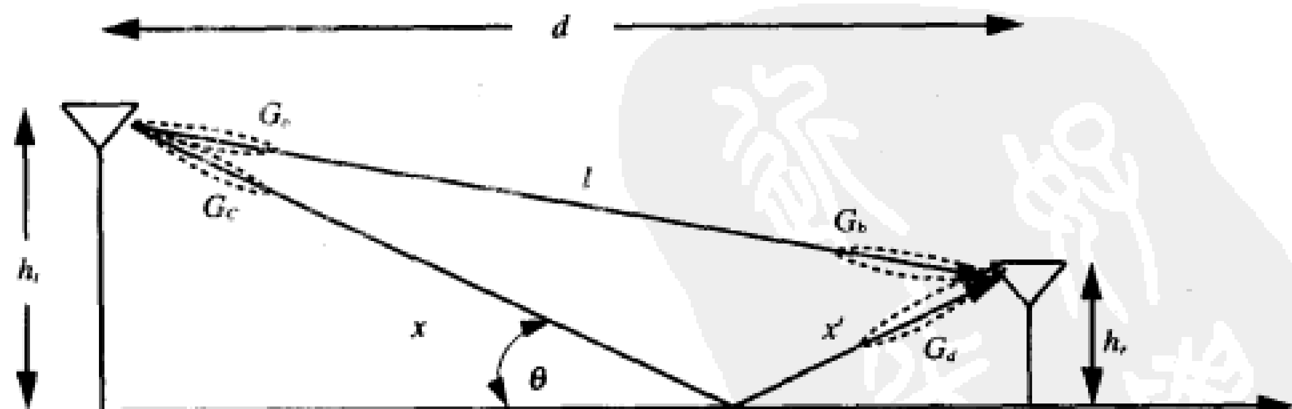


图2-4 两径模型

如果发射信号相对于时延扩展是窄带的（$\tau \ll B_u^{-1}$），那么$u(t)\approx u(t-\tau)$。于是，两径模型下，窄带信号的接收功率为

$$P_r=P_t\left[\frac{\lambda}{4\pi}\right]^2\left|\frac{\sqrt{G_l}}{l}+\frac{R\sqrt{G_r}\,\mathrm{e}^{-\mathrm{j}\Delta\phi}}{x+x'}\right|^2 \tag{2-12}$$

1. 当天线架高超过几个波长时，这是一个合理的近似。

其中$\Delta\phi=2\pi(x+x'-l)/\lambda$是直射信号和反射信号的相位差。如文献[15]所示，式（2-12）与经验数据相近。若d表示收发天线的水平距离、h_t表示发送天线高度、h_r表示接收天线高度，那么由几何关系可得

$$x+x'-l=\sqrt{(h_t+h_r)^2+d^2}-\sqrt{(h_t-h_r)^2+d^2} \tag{2-13}$$

当$d>>h_t+h_r$时，由泰勒级数近似可得

$$\Delta\phi=\frac{2\pi(x+x'-l)}{\lambda}\approx\frac{4\pi h_t h_r}{\lambda d} \tag{2-14}$$

地面反射系数由下式给出[6, 16]：

$$R=\frac{\sin\theta-Z}{\sin\theta+Z} \tag{2-15}$$

其中

$$Z=\begin{cases}\sqrt{\varepsilon_r-\cos^2\theta}/\varepsilon_r & \text{垂直极化}\\ \sqrt{\varepsilon_r-\cos^2\theta} & \text{水平极化}\end{cases} \tag{2-16}$$

ε_r是大地的介电常数。陆地及道路的介电常数近似于纯绝缘体，是一个约等于15的实数。

35

从图2-4和式（2-15）可见，当d充分大时，$x+x'\approx l\approx d$、$\theta\approx 0$、$G_l=G_r$、$R\approx-1$。把这些近似关系代入式（2-12），则接收功率近似为

$$P_r\approx\left[\frac{\lambda\sqrt{G_l}}{4\pi d}\right]^2\left[\frac{4\pi h_t h_r}{\lambda d}\right]^2 P_t=\left[\frac{\sqrt{G_l}h_t h_r}{d^2}\right]^2 P_t \tag{2-17}$$

或可表示成分贝形式为

$$P_r\text{dBm}=P_t\text{dBm}+10\log_{10}(G_l)+20\log_{10}(h_t h_r)-40\log_{10}(d) \tag{2-18}$$

因此，当d充分大时，接收功率与d的四次方成反比，并且与波长λ无关。接收信号与λ无关的原因是式中的天线的接收功率并没有随频率增大而单调减小，直射路径和反射路径的叠加实际上等效形成了一个天线阵列。图2-5根据式（2-12）画出了接收功率随距离变化的曲线，图中条件是$f=900\text{MHz}$、$R=-1$、$h_t=50\text{m}$、$h_r=2\text{m}$、$G_l=1$、$G_r=1$。曲线起始处的发送功率已归一化为0dBm。这条曲线可以分成三段：距离较小（$d<h_t$）时，两个路径分量正向相加，接收功率随距离缓慢增加，此时发送机与接收机的距离为$l=\sqrt{d^2+(h_t-h_r)^2}$又因为$h_t>>h_r$，故$1/l^2\approx1/(d^2+h_t^2)$所以接收功率准确地说正比于$1/(d^2+h_t^2)$；当距离$d$大于$h_t$而小于一个临界值$d_c$时，两路电波的干涉形成了一系列极大极小值。这种情形也被称为小尺度衰落或多径衰落，下章将对此进行进一步的讨论；曲线在临界距离d_c处达到最后一个极大值，此后信号功率随d^{-4}下降。这是因为$d>d_c$时，两个路径上信号分量的相位差大约为π，只存在反向叠加。令式（2-14）中$\Delta\phi=\pi$可近似得到临界距离为$d_c=4h_t h_r/\lambda$，此值已标示在图中。将两径模型中局部的极值进行平均，就可以近似得到功率随距离增加而下降的趋势，为一个三段折线模型，如图2-5所示（图中为了清楚起见，有意稍微偏离了曲线）。在第一段中，功率是恒定的并正比于$1/h_t^2$；对于距离介于h_t和d_c之间的第二段，每10倍距离功率下降20dB；对于大于d_c的距离，每10倍距离功率下降40dB。

36

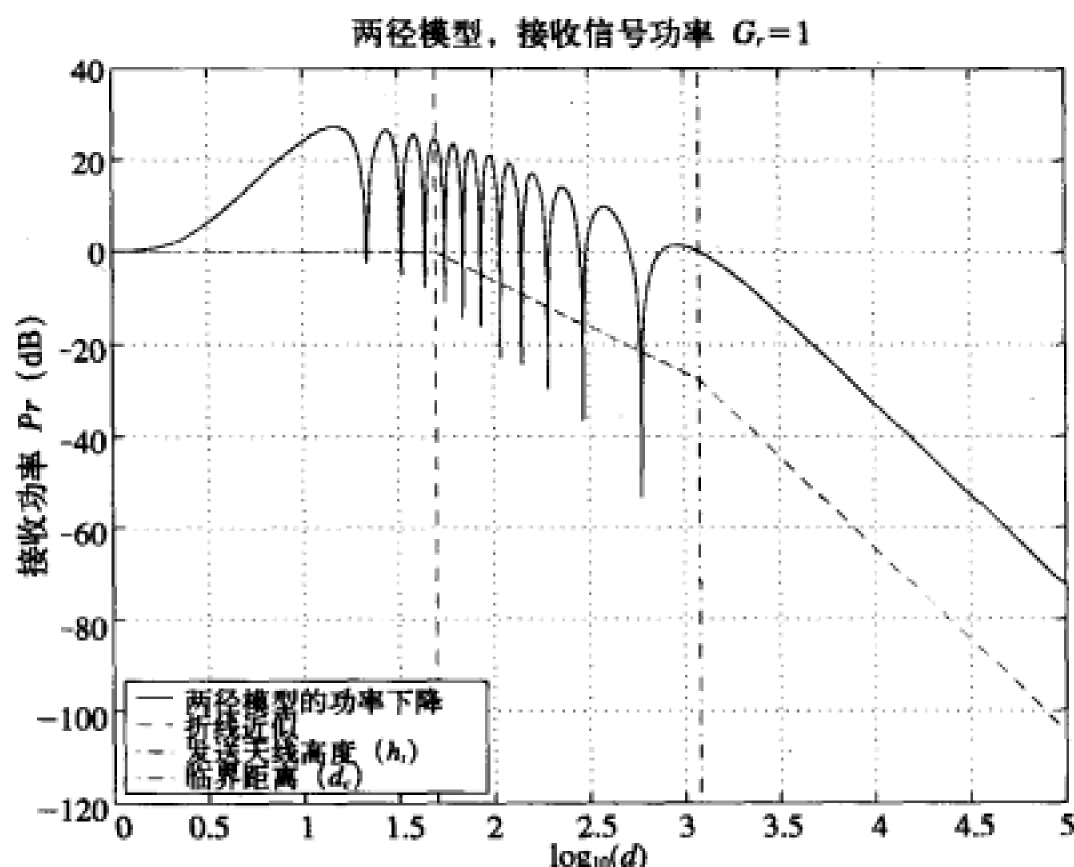


图2-5 两径模型中接收功率与距离的关系

临界距离d_c可用于系统设计。例如，若蜂窝系统的电波传播符合两径模型，那么把临界距离设计为小区半径，将使外小区干扰的路径损耗比本小区有用信号的路径损耗大得多。不过从图2-5和下面的例子来看，小区半径设为d_c有些大。出于增加容量和降低发射功率的考虑，一般采用半径比较小的小区，比d_c小得多。若这些较小的小区服从两径传播模型，那么有用功率和干扰功率都随着距离的平方下降。蜂窝系统中的电波传播很少符合两径模型，因为其他所有方向上的多径分量相互抵消，只剩地面反射波和直射波的可能性极小。

例2.2：请计算城市微小区（$h_t=10$m、$h_r=3$m）和室内微小区（$h_t=3$m、$h_r=2$m）两径模型的临界距离，发射频率f_c为2GHz。

解：由$d_c=4h_th_r/\lambda$，可得出d_c对于城市微小区为800m，对于室内微小区为160m。800m的半径对于城市微小区系统来说有点大，现在的城市微小区为了保证容量，半径一般都取在100m左右。如果我们在这些系统参数下采用了800m半径的小区，那么小区内的信号功率将随d^2下降，而邻小区的干扰随d^4衰减，干扰就可以大大降低。类似的，160m的小区半径对于室内系统也有些过大，这将要求信号能够穿透许多墙壁。因此室内系统的小区半径一般较小，约为10m~20m。

2.4.2 十径模型（介电峡谷）

下面我们介绍Amitay[9]提出的一种城市微小区模型。这种模型将环境假定为一个方

方正正的城市[1]，街道两边是建筑物，发射天线和接收天线的高度接近于地面。两旁排列着建筑物的街道对无线电信号相当于一个介电峡谷。理论上，经由临街的建筑物反射至接收机的射束有无限个，另外还可能有从后街反射来的信号。由于每次反射都会衰减信号的能量，因而经过三次以上反射的路径可以忽略不计。当街道布局较为平直时，后街的反射信号也可以忽略。实验数据表明，有十个反射路径的模型与介电峡谷中信号的传播极为相似[9]。十径模型包括各种一次、两次和三次反射信号，具体有直射路径（LOS）、地面反射路径（GR）、一次墙面反射路径（SW）、两次墙面反射路径（DW）、三次墙面反射（TW）路径、墙地反射路径（WG）和地墙反射路径（GW）。对于街道的两边，每种墙面反射都有两条路径。街道的两边各有一条路径，图2-6为十径模型的俯视图。

37

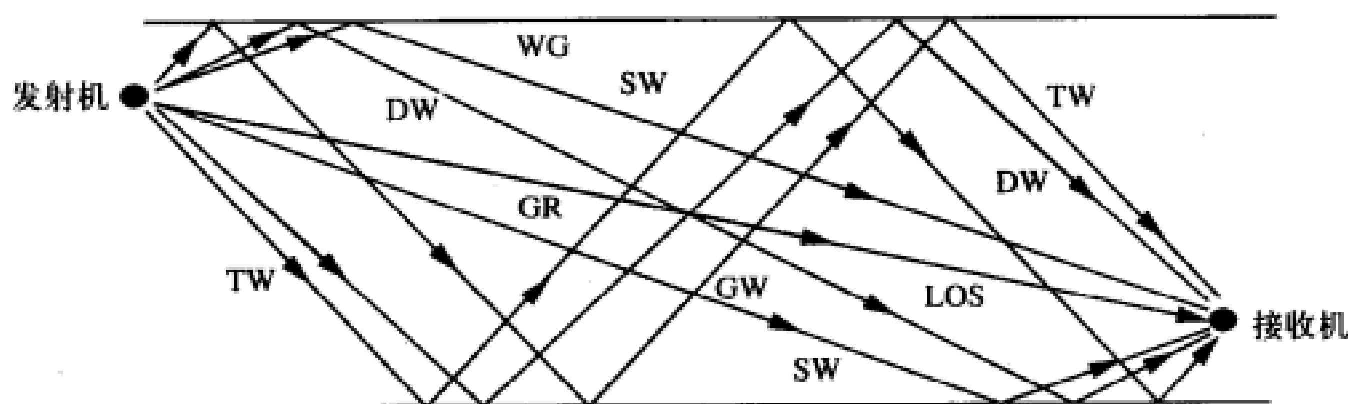


图2-6　十径模型的俯视图

十径模型下的接收信号为

$$r_{10\text{-ray}}(t)=\mathrm{Re}\left\{\frac{\lambda}{4\pi}\left[\frac{\sqrt{G_l}u(t)\mathrm{e}^{-\mathrm{j}2\pi l/\lambda}}{l}+\sum_{i=1}^{9}\frac{R_i\sqrt{G_{x_i}}u(t-\tau_i)\mathrm{e}^{-\mathrm{j}2\pi x_i/\lambda}}{x_i}\right]\mathrm{e}^{\mathrm{j}2\pi f_c t}\right\} \tag{2-19}$$

式中x_i代表第i个反射路径的长度，$\tau_i=(x_i-l)/c$，$\sqrt{G_{x_i}}$是第i条路径的发射天线增益和接收天线增益之积。对于单反射路径，R_i是式（2-15）定义的反射系数。对于多反射路径，R_i为每次反射的反射系数之积。式（2-19）中用的介电常数近似等于大地介电常数，即用$\varepsilon_r=15$来计算所有的R_i。如果又假设是窄带模型，则对于所有的i有$u(t)\approx u(t-\tau_i)$，此时对应于式（2-19）的接收功率为

$$P_r=P_t\left[\frac{\lambda}{4\pi}\right]^2\left|\frac{\sqrt{G_l}}{l}+\sum_{i=1}^{9}\frac{R_i\sqrt{G_{x_i}}\mathrm{e}^{-\mathrm{j}\Delta\phi_i}}{x_i}\right|^2 \tag{2-20}$$

其中$\Delta\phi_i=2\pi(x_i-l)/\lambda$。

无论发射天线是高于或低于建筑物，由式（2-20）的十径模型得到的接收功率随距离的平方下降，其趋势和城市实测数据一样[15, 17~18]，即使当距离很大时也如此。功率下降指数对天线高度并不敏感是因为在十径模型中，直射波和地面反射波构成的叠加（按d^{-4}衰减的两径模型）所占的比重小，多径的叠加起着主要的作用，它们按d^{-2}下降。其他实验研究[19~21]的结论是功率随$d^{-\gamma}$衰减，γ介于2和6之间。

1. 指城市在一个平面上，所有街道都是直的，街道与街道相互垂直，如曼哈顿街区。

2.4.3 通用路径跟踪

对于任意建筑物布局和任意天线位置的情形，可用通用路径跟踪（general ray tracing，GRT）来预测场强和时延扩展[22-24]。它需要准确的建筑物数据（高度、位置和介电特性），以及收发天线相对于建筑物的位置。这些信息和特定的站址有相关，所以GRT模型不用于研究一般性的系统性能和网络规划问题，而是为了解释城市电波传播的基本机理，并可针对具体的环境的发射机接收机布点，得到信号的时延和强度。

38

GRT模型用几何光学的方法跟踪直射信号分量、反射信号分量以及由建筑物绕射和漫散射的信号分量。对于给定的接收机位置，GRT根据具体的建筑物位置和介电性质导出任意个多径分量的强度。由于绕射和散射损耗较大，直射径和反射径构成了接收信号的主要成分。但对于靠近散射体和衍射体的区域，直射和反射分量可能被阻挡，此时其余多径分量将起主导作用。

直射径和反射径的传播模型已经在前一节进行了介绍。绕射就是发送信号绕过发射机和接收机之间的障碍物而到达接收机，如图2-7所示。地球弯曲的表面、多山地形、不规则的地形、建筑物的边缘和收发视距上的障碍物等多种情况都可能发生绕射[1, 5, 16]。绕射特性可通过几何绕射理论（geometrical theory of diffraction，GTD）[25]精确计算，但其复杂性妨碍了它在无线信道建模中的应用。楔形绕射把产生绕射的物体假设为楔形，可大大简化几何绕射理论。该模型已用于分析街道拐角处的绕射现象，对于一定的入射角，这种绕射可以导致超过100dB的路径损耗[11, 24, 26-27]。楔形绕射虽然简化了GTD，但仍然需要通过数值方法求解路径损耗[25, 28]，所以没有被广泛使用。**菲涅尔刃形绕射模型**（Fresnel knife-edge diffraction model）则因其简单性而成为最常用的一种绕射模型。该模型示于图2-7。它假设绕射体非常薄，一般的山地、崎岖地形或者楔形绕射体都不符合这个假设。它也没有考虑极化、导电率和表面粗糙程度等绕射参数，这些因素会导致其结果不够准确[26]。图2-7中绕射信号的传播距离是$(d+d')$，对应的相移为$\phi=2\pi(d+d')/\lambda$。由图2-7的几何关系可得，若h远小于d和d'，则信号的绕射径相对于直射径的行程差约为

$$\Delta d=\frac{h^2}{2}\frac{d+d'}{dd'}$$

相对于直射路径的相位移大约为

$$\Delta\phi=\frac{2\pi\Delta d}{\lambda}\approx\frac{\pi}{2}v^2 \tag{2-21}$$

39

其中

$$v=h\sqrt{\frac{2(d+d')}{\lambda dd'}} \tag{2-22}$$

被称为**菲涅尔-基尔霍夫**（Fresnel-Kirchhoff）绕射参数。刃形绕射带来的路径损耗一般是v的函数。计算绕射损耗相当复杂，要用到惠更斯原理、菲涅尔区和复杂的菲涅尔积分[1]，结果一般无闭式解。李建业（William CY Lee）在文献[16]的第2章中给出了一个近似式，可计算相对于直射径的刃形绕射损耗（分贝值）

$$L(v)\mathrm{dB}=\begin{cases}20\log_{10}[0.5-0.62v] & -0.8v\leqslant v<0\\ 20\log_{10}[0.5\mathrm{e}^{-0.95v}] & 0\leqslant v<1\\ 20\log_{10}[0.4-\sqrt{0.1184-(0.38-0.1v)^2}\,] & 1\leqslant v\leqslant 2.4\\ 20\log_{10}[0.225/v] & v>2.4\end{cases} \tag{2-23}$$

文献[29]中也有类似的近似式。由刃形绕射模型可得到绕射接收信号为

$$r(t)=\mathrm{Re}\left\{\lambda/(4\pi(d+d'))L(v)\sqrt{G_d}u(t-\tau)\mathrm{e}^{-\mathrm{j}2\pi(d+d')/\lambda}\mathrm{e}^{\mathrm{j}2\pi f_c t}\right\} \tag{2-24}$$

式中 $\sqrt{G_d}$ 为天线增益，$\tau=\Delta d/c$为绕射路径相对于直射路径的时延。

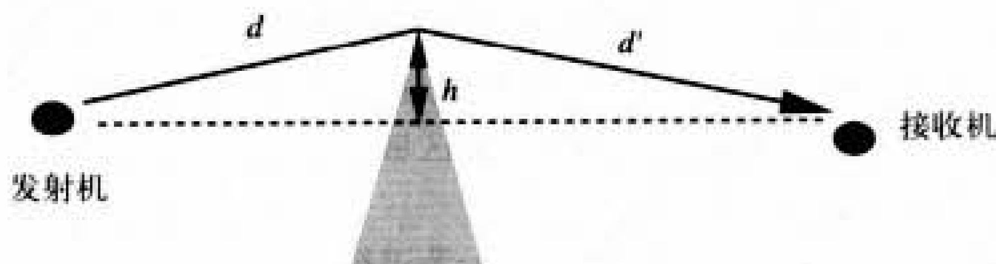


图2-7 刃形绕射

除上述外，还可能有多次绕射的路径，以及同时经过了反射和绕射的路径。对于反射和绕射的各种组合情况也有相应的模型[30]，这些信号分量的衰减一般都非常大，以至于相对于噪声可以忽略。绕射模型也可以针对特定的环境建立，例如Walfisch和Bertoni给出了针对蜂窝系统中屋顶以及建筑物的绕射模型[31]。

散射波由示于图2-8中的s和s'两段形成。电磁波经过散射后又有一次扩散损耗，所以散射损耗正比于s和s'之积。接收到的散射信号由双基站雷达方程[32]给出：

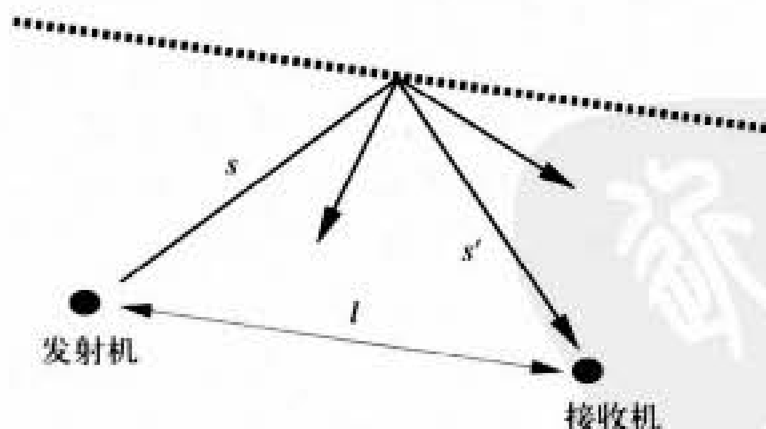


图2-8 散射

$$r(t)=\mathrm{Re}\left\{u(t-\tau)\frac{\lambda\sqrt{G_s}\sigma\mathrm{e}^{-\mathrm{j}2\pi(s+s')/\lambda}}{(4\pi)^{3/2}ss'}\mathrm{e}^{\mathrm{j}2\pi f_c t}\right\} \tag{2-25}$$

式中$\tau=(s+s'-l)/c$是散射波的相对时延；σ（平方米）为散射物的雷达截面积，取决于散射体的粗糙程度、大小和形状，$\sqrt{G_s}$是天线增益。该模型假定信号按自由空间模型从发射机传播到散射体，在散射体处再以散射体接收功率的σ倍向外辐射。由式（2-25），散射引起的路径损耗为

$$P_r\text{dBm}=P_t\text{dBm}+10\log_{10}(G_s)+20\log_{10}(\lambda)+10\log_{10}(\sigma)-30\log_{10}(4\pi)-20\log_{10}(s)-20\log_{10}(s') \tag{2-26}$$

文献[33]对多个城市的不同建筑物给出了$10\log_{10}\sigma$的经验值。其结果表明，$10\log_{10}\sigma$的范围为-4.5dBm^2到55.7dBm^2，此处dBm^2为以平方米为单位的σ的分贝值。

接收到的信号是所有多径分量的叠加。如果有一条直射路径、N_r条反射路径、N_d条绕射路径和N_s条散射路径，则总的接收信号可表示为

$$\begin{aligned}r_{\text{total}}(t)=\text{Re}\Bigg\{\left[\frac{\lambda}{4\pi}\right]\Bigg[&\frac{\sqrt{G_l}u(t)e^{j2\pi l/\lambda}}{l}+\sum_{i=1}^{N_r}\frac{R_{x_i}\sqrt{G_{x_i}}u(t-\tau_i)e^{-j2\pi x_i/\lambda}}{x_i}\\&+\sum_{j=1}^{N_d}\frac{4\pi}{\lambda}L_j(v)\sqrt{G_{d_j}}u(t-\tau_j)e^{-j2\pi(d_j+d_j')/\lambda}\\&+\sum_{k=1}^{N_s}\frac{\sqrt{G_{s_k}\sigma_k}u(t-\tau_k)e^{-j2\pi(s_k+s_k')/\lambda}}{\sqrt{4\pi s_k s_k'}}\Bigg]e^{j2\pi f_c t}\Bigg\}\end{aligned} \tag{2-27}$$

式中τ_i、τ_j和τ_k分别代表反射、绕射和散射路径相对于直射路径的时延。由式（2-27）可求得$r_{\text{total}}(t)$的接收功率P_r和相应的路径损耗P_r/P_t。

如果这些多径分量中的任何一径被建筑物或其他物体阻挡，就会有一个附加的衰减因子乘到式（2-27）的对应项中。根据阻挡物材质和阻挡深度的不同，这个衰减因子会有很大的差异[5, 34]。2.7节将讨论这种衰减造成的随机损耗的模型。

41

2.4.4 本地平均接收功率

所有射线跟踪模型中路径损耗的计算都是在发射机和接收机位置固定的情况下进行的。射线跟踪中可用所有路径的幅度平方和求出接收机位置附近的*本地平均接收功率*（local mean received power）$\overline{P}_r$。这种方法平均了给定位置处因为相位变化引起的接收功率在附近空间上的变化。本地平均接收功率能较好地反映链路质量，经常用在蜂窝系统的功率控制和切换中[35]。

2.5 经验路径损耗模型

大多数移动通信系统运行在复杂的传播环境中，这些传播环境难以用自由空间路径损耗或射线跟踪法精确建模。多年来，人们针对不同的环境又提出了许多基于实测数据的经验模型，包括城市宏小区、城市微小区等，最近还有针对室内环境的模型（见文献[5]的第3章）。先针对特定的环境按不同的距离和频率取得测量数据，再用这些数据建模。所建立的模型并不局限于取得数据的那个环境，而是要广泛用到其他通用的传播环境中。这样做虽然有准确性的问题，但许多无线系统都用这样的经验模型作为系统性能分析的基础。我们下面将先介绍城市宏小区的常见模型，然后再介绍新近出现的室外微小区和室内传播的模型。

在前几节提到的解析模型中，路径损耗P_r/P_t作为距离的函数有明确的含义。但对于实测数据来说，P_r/P_t还包括路径损耗、阴影和多径的影响。为此，实测时为了得到路径

损耗，一般是把附近几个波长范围内的测量数据进行平均，以消除多径的影响。这样得到的平均路径损耗叫做距离d处的*本地平均损耗*（local mean attenuation，LMA）。由于自由空间损耗和信号阻挡，LMA一般随距离d的增加而下降。对于城市等具体环境，本地平均损耗和测量时发射机和接收机的具体位置有关。为使结果具有一般性，一般要测遍所在环境，并可能要对多个特性类似的环境进行测量，再将特定环境（如城市、郊区和办公室等）下给定距离d处的测量结果进行平均，得到*经验路径损耗*（empirical path loss）$P_L(d)$。例如，在纽约、旧金山、芝加哥市中心测量本地平均损耗，再将这些测量结果做平均就得到一般矩形格状街区的经验路径损耗。本节所给出的经验路径损耗模型都是这样得到的。

42

2.5.1 奥村模型（Okumura Model）

奥村模型[36]是城市宏小区中信号预测最常用的模型之一。该模型适用的距离范围是1km~100km、频率范围是150MHz~1 500MHz。奥村对东京地区基站到移动台的信号传播衰减进行了大量的测量，用一系列曲线的形式给出了不规则地形条件下，相对于自由空间传播的损耗中值。测量中基站高度为30m~100m，此高度范围的上界高于现用基站的典型高度。奥村模型的经验路径损耗公式可以表示为

$$P_L(d)\text{dB} = L(f_c, d) + A_\mu(f_c, d) - G(h_t) - G(h_r) - G_{\text{AREA}} \tag{2-28}$$

式中$L(f_c, d)$是传播距离为d、载频为f_c时的自由空间路径损耗，$A_\mu(f_c, d)$是针对所有环境所附加的衰减中值，$G(h_t)$为基站天线高度增益因子，$G(h_r)$为移动台天线高度增益因子，G_{AREA}是体现传播环境的增益因子。$A_\mu(f_c, d)$和G_{AREA}的值可从奥村的经验曲线图得到[36, 5]。奥村给出的计算$G(h_t)$、$G(h_r)$的经验公式为：

$$G(h_t) = 20\log_{10}(h_t/200) \quad 30\text{m} < h_t < 1000\text{m} \tag{2-29}$$

$$G(h_r) = \begin{cases} 10\log_{10}(h_r/3) & h_r \leqslant 3\text{m} \\ 20\log_{10}(h_r/3) & 3\text{m} < h_r < 10\text{m} \end{cases} \tag{2-30}$$

为了提高模型的准确度，文献[36]还给出不同地形的校正因子。奥村模型所预测的路径损耗和用实测数据建模所得的路径损耗相比，误差的标准差约为10dB~14dB。

2.5.2 哈塔模型（Hata Model）

哈塔模型[37]把奥村用曲线图表示的路径损耗数据拟合为经验公式，适用频率范围也基本是150MHz~1500MHz。该模型闭式的公式简化了路径损耗的计算，不需要通过查经验曲线来确定相关参数。城市地区哈塔模型经验路径损耗的标准公式为

$$\begin{aligned} P_{L,\text{urban}}(d)\text{dB} = 69.55 + 26.16\log_{10}(f_c) - 13.82\log_{10}(h_t) - a(h_r) \\ + (44.9 - 6.55\log_{10}(h_t))\log_{10}(d) \end{aligned} \tag{2-31}$$

式中参数的含义和奥村模型一样，$a(h_r)$为移动台天线高度校正因子。对于中小城市，该因子由下式给出[37, 5]

$$a(h_r) = (1.1\log_{10}(f_c) - 0.7)h_r - (1.56\log_{10}(f_c) - 0.8)\text{dB}$$

对于较大的城市，当$f_c>300$MHz时

$$a(h_r)=3.2(\log_{10}(11.75h_r))^2-4.97\text{dB}$$

城市哈塔模型经校正后也可用于郊区和乡村，其式分别为

$$P_{L,\text{suburban}}(d)\text{dB}=P_{L,\text{urban}}(d)\text{dB}-2[\log_{10}(f_c/28)]^2-5.4 \tag{2-32}$$

$$P_{L,\text{rural}}(d)\text{dB}=P_{L,\text{urban}}(d)\text{dB}-4.78[\log_{10}(f_c)]^2+18.33\log_{10}(f_c)-K \tag{2-33}$$

其中K的变化范围为35.94（乡村）到40.94（沙漠）。与奥村模型不同，哈塔模型不再对特定的传输路径进行校正。当$d>1$km时，哈塔模型是奥村模型的很好近似。哈塔模型对于第一代蜂窝系统来说是一种非常好的模型，但它不能很好体现目前使用较小小区和较高频率的蜂窝系统中的传播特性。哈塔模型也不能反映室内传播特性。

2.5.3 哈塔模型的COST231扩展

欧洲科技合作组织（European cooperation for scientific and technical research，EURO-COST）将哈塔模型扩展到2GHz，成为[38]

$$\begin{aligned}P_{L,\text{urban}}(d)\text{dB}=&46.3+33.9\log_{10}(f_c)-13.82\log_{10}(h_t)-a(h_r)\\&+(44.9-6.55\log_{10}(h_t))\log_{10}(d)+C_M\end{aligned} \tag{2-34}$$

式中$a(h_r)$和哈塔模型里的校正因子相同。对于中等城市和郊区C_M取0dB，对于大型城市取3dB。这个模型被称作哈塔模型的COST231扩展，适用范围限定于：1.5GHz<f_c<2GHz、30m<h_t<200m、1m<h_r<10m、1km<d<20km。

2.5.4 折线（多斜率）模型

室外微小区和室内信道中常用折线模型来表示经验路径损耗分贝值和距离对数的关系，如图2-9所示。图中的点代表假想的测量结果，折线是对这些测量值的近似。N段折线模型需给定（$N-1$）个转折点$d_1,\cdots,d_{N-1}$和每段的斜率$s_1,\cdots,s_N$。建模时可先通过多种方法确定转折点的数量和位置。然后通过线性回归得到每段的斜率。文献[39]用折线模型来反映室外信道的路径损耗，而文献[40]用折线模型反映室内信道的路径损耗。

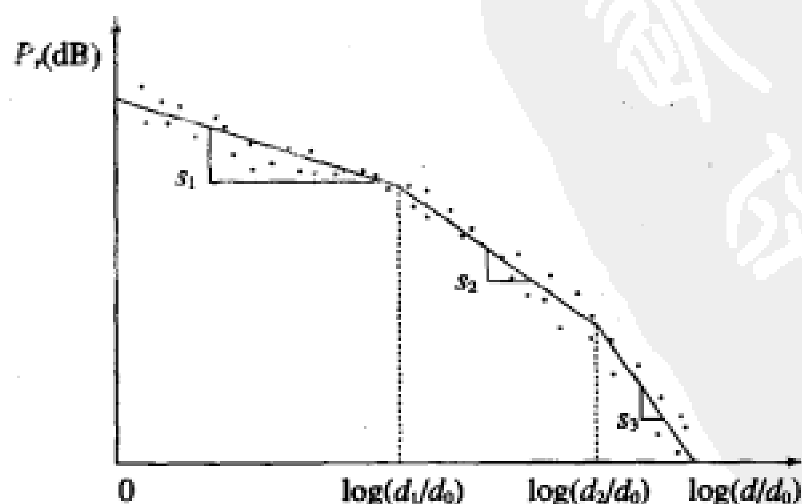


图2-9 路径损耗的折线模型

折线模型的一个特例是双斜率模型，其参数包括：确定的路径损耗因子K、参考距离d_0和临界距离d_c之间的路径损耗指数γ_1、距离大于d_c时的路径损耗指数γ_2：

$$P_r(d)\text{dB}=\begin{cases}P_t+K-10\gamma_1\log_{10}(d/d_0) & d_0\leqslant d\leqslant d_c\\ P_t+K-10\gamma_1\log_{10}(d_c/d_0)-10\gamma_2\log_{10}(d/d_c) & d>d_c\end{cases}\quad(2\text{-}35)$$

一般用回归的方法从经验数据中拟合出路径损耗指数、K和d_c[41~42]。2.4.1节中的两径模型在$d>h_t$时可近似为双斜率模型，临界距离d_c处为转折点，两个斜率分别是$s_1=20$dB/十倍程和$s_2=40$dB/十倍程。

44

分段函数形式的双斜率模型还可近似为[19, 43]：

$$P_r=\frac{P_tK}{L(d)}\quad(2\text{-}36)$$

其中

$$L(d)\triangleq\left[\frac{d}{d_0}\right]^{\gamma_1}\left(1+\left(\frac{d}{d_c}\right)^{(\gamma_2-\gamma_1)q}\right)^{1/q}\quad(2\text{-}37)$$

参数q反映转折点附近路径损耗变化的平滑程度。这一模型也可扩展到多段[39]。

2.5.5 室内衰减因子

室内传播因具体环境的不同有很大的差异，有许多因素对室内路径损耗有显著影响，包括墙壁和地板层的材质、室内布局、走廊、窗户、开阔区域、阻挡物的位置和材料、每个房间的大小和地板层数等。因此，很难建立出能准确计算特定室内环境中的路径损耗的通用模型。

室内路径损耗模型必须能够很好地反映来自地板层和隔墙的信号衰减。对大量不同特性的建筑物和大范围的信号频率进行的测量表明，信号第一次穿过地板时衰减最大，以后穿过时衰减量递减。具体来说，文献[44~47]的测量表明，对于900MHz，发射机和接收机的位置相隔一层时，衰减为10dB~20dB；接下来相隔的三层中每层衰减为6dB~10dB；再接下来的四层中，每层衰减只有少许几个分贝。对于更高的频率，每层的衰减一般会更大[45, 48]。每层的衰减量随着层数的增多而下降，这是由于楼边的散射和附近建筑物的反射造成的。隔墙损耗因材质和介电性质的不同而有很大差别。不同频率、不同隔墙类型时损耗的测量数据见文献[5、44、49~51]，表2-1是其中的一些例子，对应的频率是900MHz~1300MHz。即使对于相同的频率和相同类型的隔墙，不同研究者得出损耗差别也非常大，因此很难得到一般化的结论。

45

表2-1　典型隔墙损耗

隔墙类型	隔墙损耗（dB）	隔墙类型	隔墙损耗（dB）
布料	1.4	混凝土墙	13
双层石膏墙板	3.4	铝墙板	20.4
绝缘箔型	3.9	全金属型	26

可将地板层损耗和隔墙损耗的经验数据加到路径损耗$P_L(d)$的解析模型的或经验模型中：

$$P_r\text{dBm} = P_t\text{dBm} - P_L(d) - \sum_{i=1}^{N_f}\text{FAF}_i - \sum_{i=1}^{N_P}\text{PAF}_i \tag{2-38}$$

式中FAF_i、PAF_i分别代表信号所穿过第i个地板层或隔墙的衰减因子，N_f、N_P分别是地板层数和隔墙数。

当发射机在建筑物外时，对于室内系统还有一个重要参数是建筑物穿透损耗。测量表明，建筑物穿透损耗与频率、高度和建筑物材料有关。对于900MHz~2GHz的频率范围，一层的建筑物穿透损耗的典型值为8dB~20dB[1, 52-53]。穿透损耗随频率的增加略有下降。此外，每增高一层，穿透损耗下降1.4dB。这是由于楼层越高时，建筑物密集程度降低，出现直射径的可能性增大。建筑物内窗户的类型和数量也对穿透损耗有重要的影响[54]。在窗户后测得的穿透损耗比在外墙后测到的值小6dB。平板玻璃的穿透损耗约为6dB，铅化玻璃的损耗在3dB~30dB之间。

2.6 简化的路径损耗模型

信号传播的复杂性使我们很难用单一模型来精确反映各种传播环境下的路径损耗。对于比较严格的要求或对于基站和接入点的最佳选址问题，我们可以采用复杂的解析模型或通过实测来建立精确模型。但如果只是为了对不同的系统设计进行一般性的优劣分析，则可以不用复杂模型，用一个能反映信号传播主要特性的简单模型即可。对实际信道来说，这样做当然只是一种近似。因此，系统设计中经常使用下面的简化路径损耗模型：

46

$$P_r = P_t K\left[\frac{d_0}{d}\right]^\gamma \tag{2-39}$$

对应分贝值为

$$P_r\text{dBm} = P_t\text{dBm} + K\text{dB} - 10\gamma\log_{10}\left[\frac{d}{d_0}\right] \tag{2-40}$$

式中，K是一个依赖于天线特性和平均信道损耗的常系数、d_0为天线远场的参考距离、γ为路径损耗指数。可选择合适的K、d_0和γ来近似解析模型或经验模型。例如，自由空间模型、两径模型、哈塔模型和COST扩展的哈塔模型都具有式（2-40）的形式。由于天线近场区存在散射现象，式（2-40）一般只适用于发送距离$d>d_0$时，d_0值对于室内一般为1m~10m、对于室外一般为10m~100m。

用简化模型近似实测数据时，常把$K<1$取为全向天线在d_0处的自由空间路径增益：

$$K\text{dB} = 20\log_{10}\frac{\lambda}{4\pi d_0} \tag{2-41}$$

文献[41]在传送距离100m处实测的自由空间路径损耗与此相符。K值也可以用传播距离d_0处的实测数据来确定，并可以通过最小化模型和实测数据之间的均方误差（mean square error，MSE）来优化（并一同优化γ）[41]。γ的值依赖于传播环境，信号传播近似于自由空间模型时γ取2，近似于两径模型时γ取4。对于更复杂的环境，可对实测数据进行最小均方误差（minimum mean square error，MMSE）拟合得到γ的取值，如例2.3。也可以从考虑了频率和天线高度的经验模型中得到γ[41]。表2-2列出了不同环境下的γ值（数

据来自文献[5]、[33]、[41]、[44]、[46~47]、[52]、[55]）。路径损耗指数随频率升高趋向于增大[46], [51~52], [56]，随天线增高趋于下降[41]。需注意的是，室内传播环境中，地板层、阻挡物、隔墙等因素的具体差别可能很大，造成路径损耗指数的取值范围较大，见2.5.5节。

47

表2-2　路径损耗指数的典型值

环　境	γ的范围
城市宏小区	3.7～6.5
城市微小区	2.7～3.5
写字楼（同层）	1.6～3.5
写字楼（异层）	2～6
商店	1.8～2.2
工厂	1.6～3.3
家居	3

例2.3：表2-3中给出了900MHz室内系统的一组P_r/P_t测量数据。请找出能使简化模型式（2-40）和实测数据之间的均方误差最小的路径损耗指数γ值。假定$d_0=1\text{m}$，K由距离d_0处的自由空间路径损耗公式确定。注意本题中均方误差是关于分贝值的，不是真值的均方误差，这样的处理一般更准确一些。假设发射功率为1mW（0dBm），请用所得到的简化模型计算出距离150m处的接收功率。

表2-3　路径损耗测量值

离开发射机的距离	$M=P_r/P_t$
10m	−70dB
20m	−75dB
50m	−90dB
100m	−110dB
300m	−125dB

解：分贝测量值下的最小均方误差为

$$F(\gamma)=\sum_{i=1}^{5}\left[M_{\text{measured}}(d_i)-M_{\text{model}}(d_i)\right]^2$$

其中的$M_{\text{measured}}(d_i)$是表2-3中所列的距离d_i处的路径损耗测量值，$M_{\text{model}}(d_i)=K-10\gamma\log_{10}(d_i)$是用式（2-40）计算出的距离$d_i$处的路径损耗值。用自由空间损耗公式可以得出$K=20\log_{10}(0.3333/(4\pi))=-31.54\text{dB}$，因此

$$\begin{aligned}F(\gamma)&=(-70+31.54+10\gamma)^2+(-75+31.54+13.01\gamma)^2\\&\quad+(-90+31.54+16.99\gamma)^2+(-110+31.54+20\gamma)^2\\&\quad+(-125+31.54+24.77\gamma)^2\\&=21676.3-11654.9\gamma+1571.47\gamma^2\end{aligned}\tag{2-42}$$

$F(\gamma)$对γ求导，令求导后的结果为零得

$$\frac{\partial F(\gamma)}{\partial\gamma}=-11654.9+3142.94\gamma=0\Rightarrow\gamma=3.71$$

将$K=-31.54$、$\gamma=3.71$、$P_t=0$dBm代入简化路径损耗模型，得到距离150m处的接收功率$P_r=P_t+K-10\gamma\log_{10}(d/d_0)=0-31.54-10\times3.71\log_{10}(150)=-112.27$dBm。测量值和简化路径损耗模型之间明显有偏差，这种偏差可归到阴影衰落中，2.7节将对此继续介绍。

2.7 阴影衰落

信号在无线信道传播过程中遇到的障碍物会使信号发生随机变化，从而造成给定距离处接收信号功率的随机变化，反射面和散射体的变化也会造成接收功率的随机变化。因此，需要建立一个模型来描述这些因素造成的信号随机衰减。造成信号随机衰减的因素，包括障碍物的位置、大小和介电特性及反射面和散射体的变化情况一般是未知的，因此只能用统计模型来表征这种随机衰减。最常用的模型是对数正态阴影模型，它已经被实测数据证实[41, 57]，可以精确地建模室外和室内无线传播环境中接收功率的变化。

48

对数正态阴影模型把发射和接收功率的比值$\Psi=P_t/P_r$假设为一个对数正态分布的随机变量，即

$$p(\psi)\frac{\xi}{\sqrt{2\pi}\sigma_{\psi_{dB}}\psi}\exp\left[-\frac{(10\log_{10}\psi-\mu_{\psi_{dB}})^2}{2\sigma_{\psi_{dB}}^2}\right]\qquad \psi>0 \tag{2-43}$$

式中$\xi=10/\ln10$、$\mu_{\psi_{dB}}$是以dB为单位的$\Psi_{dB}=10\log_{10}(\Psi)$的均值、$\sigma_{\psi_{dB}}$是$\Psi_{dB}$的标准差（单位也是dB）。均值可以用解析模型或者实测值确定。实测时，由于经验路径损耗的测量已经包括了对阴影衰落的平均，所以$\mu_{\psi_{dB}}$等于路径损耗。对于解析模型，$\mu_{\psi_{dB}}$必须结合考虑障碍物造成的平均衰减和路径损耗（例如从自由空间模型或射线跟踪模型中获得的路径损耗）。也可以把路径损耗从阴影衰落中分离出来单独处理，如2.8节所述。服从对数正态分布的随机变量称为对数正态随机变量（log-normal random variable）。如果Ψ为对数正态分布，那么接收功率和接收信噪比也是对数正态分布的，因为这两个量只是Ψ乘上了一个常系数。对数正态分布的接收信噪比的均值和标准差的单位也是dB，而对数正态接收功率有功率量纲，其均值和标准差的单位是dBm或dBW，而不是dB。路径损耗真值Ψ的平均值可从式（2-43）求得为

$$\mu_\psi=E[\psi]=\exp\left[\frac{\mu_{\psi_{dB}}}{\xi}+\frac{\sigma_{\psi_{dB}}^2}{2\xi^2}\right] \tag{2-44}$$

由此可得线性平均与对数平均的转换关系为

$$10\log_{10}\mu_\psi=\mu_{\psi_{dB}}+\frac{\sigma_{\psi_{dB}}^2}{2\xi} \tag{2-45}$$

对数正态阴影衰落的参数一般采用对数均值$\mu_{\psi_{dB}}$，称为平均分贝路径损耗（average dB path lass），单位是dB。经过变量代换后可以看出，Ψ的分贝值服从均值为$\mu_{\psi_{dB}}$、标准差为$\sigma_{\psi_{dB}}$的正态分布：

$$p(\psi_{dB})=\frac{1}{\sqrt{2\pi}\sigma_{\psi_{dB}}}\exp\left[-\frac{(\psi_{dB}-\mu_{\psi_{dB}})^2}{2\sigma_{\psi_{dB}}^2}\right] \tag{2-46}$$

对数正态分布由两个参数$\mu_{\psi_{dB}}$和$\sigma_{\psi_{dB}}$确定。因为$\Psi=P_t/P_r$总大于1，故$\mu_{\psi_{dB}}$总大于或等于0。注意对数正态分布（2-43）的取值范围是$0\leqslant\Psi\leqslant+\infty$。$\Psi<1$对应$P_r>P_t$，这在物理上是不可能的。不过当$\mu_{\psi_{dB}}$为很大的正值时，出现$P_r>P_t$的可能性非常小。因此，当$\mu_{\psi_{dB}}>>0$时对数正态模型能更好地反映物理实质。

用实测数据确定阴影衰落模型的均值和标准差时，会遇到这样一个问题，就是应该取真值的平均还是取分贝值的平均。具体地说，假设路径损耗测量的真值是$\{p_i\}_{i=1}^N$，那么平均路径损耗到底是按$\mu_\psi=(1/N)\sum_{i=1}^N p_i$来计算，还是应该按$\mu_{\psi_{dB}}=(1/N)\sum_{i=1}^N 10\log_{10}p_i$来计算？方差的计算也存在同样的问题。实际中常常是对测量值的分贝值取平均，来确定平均路径损耗和方差，原因有以下几点：首先，对数正态模型的数学分析是建立在分贝测量值之上的；其次，有文献表明分贝平均可以使估计误差更小[58]；其三，如2.5.4节所示，功率随距离下降的模型一般是对分贝功率测量值和对数距离的关系进行折线近似[5]。

49

多数室外信道测量表明，标准差$\sigma_{\psi_{dB}}$的范围在4dB~13dB之间[6,19,59~61]。平均值$\mu_{\psi_{dB}}$取决于路径损耗和所在区域内的建筑物属性。$\mu_{\psi_{dB}}$随距离变化，一是由于路径损耗随距离变化，二是因为距离增加时，障碍物的数量会增加，造成平均衰减增加。

当阴影衰落主要由阻挡衰减决定时，分贝平均接收功率的高斯模型可以用下面的衰减模型来分析。信号穿过宽度为d的物体时，其衰减近似为

$$s(d)=\mathrm{e}^{-\alpha d} \tag{2-47}$$

式中α是依赖于障碍物材料和介电性质的衰减常数。若第i个障碍物的衰减常数是α_i、宽度为随机值d_i，那么信号穿过该区域时的衰减为[1]，

$$s(d_i)=\mathrm{e}^{-\sum_i\alpha_i d_i} \tag{2-48}$$

如果发射机和接收机之间有多个障碍物物，那么由中心极限定理，$\sum_i\alpha_i d_i$可近似为高斯随机变量。这样，$\log s(d_i)$就是一个均值为μ，方差为σ的高斯随机变量（σ的值由传播环境决定）。

例2.4：例2.3用表2-3中的测量值拟合出了简单路径损耗模型的最佳路径损耗指数为$\gamma=3.71$。假设简化路径损耗模型取此指数，K取-31.54dB。请求出实测数据围绕平均路径损耗的对数正态阴影方差$\sigma^2_{\psi_{dB}}$。

解：对于$\gamma=3.71$的简化路径损耗模型，样值的方差为

$$\sigma^2_{\psi_{dB}}=\frac{1}{5}\sum_{i=1}^{5}[M_{\text{measured}}(d_i)-M_{\text{model}}(d_i)]^2$$

其中$M_{\text{measured}}(d_i)$是表2-3中距离d_i处的路径损耗测量值，$M_{\text{model}}(d_i)=K-37.1\log_{10}(d)$。于是

$$\begin{aligned}\sigma^2_{\psi_{dB}}&=\frac{1}{5}[(-70-31.54+37.1)^2+(-75-31.54+48.27)^2+(-90-31.54+63.03)^2\\&\quad+(-110-31.54+74.2)^2+(-125-31.54+91.90)^2]=13.29\end{aligned}$$

1. 作者对该公式进行了勘误，此处采用的是勘误后的公式。——译者注

因此，阴影衰落的标准差为$\sigma_{\psi_{dB}}=3.65$dB。注意上式方括号中的部分就是$\gamma=3.71$时例2.3中的最小均方误差公式（2-42）。

50

为了得到阴影衰落在距离上的自相关特性，人们在不同传播环境、不同频率下进行了大量测量[60, 62-65]。最常用的解析模型是Gudmundson根据实测结果提出的[60]，它将阴影衰落$\psi(d)$假设为一阶自回归过程。对于距离为δ的两个点处的阴影衰落，假设它们有相同的均值$\mu_{\psi_{dB}}$，则其协方差为

$$A(\delta)=E\left[(\psi_{dB}(d)-\mu_{\psi_{dB}})(\psi_{dB}(d+\delta)-\mu_{\psi_{dB}})\right]=\sigma_{\psi_{dB}}^2\rho_D^{\delta/D} \tag{2-49}$$

其中ρ_D是距离为D时两点的归一化协方差，只能通过实测得到，它与传播环境和载波频率有关。实测数据表明，郊区宏小区在$f_c=900$MHz、$D=100$m时$\rho_D=0.82$，城市微小区在$f_c\approx2$GHz、$D=10$m时$\rho_D=0.3$[60,64]。如置$D=X_c$处，$\rho_D=1/\mathrm{e}$，此模型可简化成不依赖于实测数据的形式：

$$A(\delta)=\sigma_{\psi_{dB}}^2\mathrm{e}^{-\delta/X_c} \tag{2-50}$$

其中的X_c叫做*去相关距离*（decorrelation distance），当距离等于此值时，协方差的最大值是1/e。X_c的大小近似等于阻挡物大小或者等于障碍物聚集体的大小。对于室外系统，X_c的典型值范围是50m~100m[63-64]。若用户以速度v移动，将$v\tau=\delta$代入式（2-49）或（2-50）可得到阴影衰落的不相关时间。文献[62]和[64]研究了相对于角度扩展的阴影自相关，它对于第10章讲述的多天线系统十分有用。

一阶自回归相关模型式（2-49）及其简化形式（2-50）易于进行分析和仿真。比如可以这样来仿真ψ_{dB}：先产生一个功率为$\sigma_{\psi_{dB}}^2$的高斯白噪声过程，然后使它通过一个响应为$\rho_D^{\delta/D}$的一阶滤波器便得到协方差为（2-49）的阴影衰落过程，或通过一个响应为$\mathrm{e}^{-\delta/X_c}$的一阶滤波器则得到协方差为式（2-50）的阴影衰落过程[60-61]。

2.8　路径损耗和阴影衰落的混合模型

将路径损耗模型和阴影衰落模型叠加在一起就可以同时反映出功率随距离的减小和阴影造成的路径损耗随机衰减。此混合模型用路径损耗模型来描述平均分贝路径损耗$\mu_{\psi_{dB}}$，再增加一个均值为0dB的阴影衰落来体现围绕路径损耗的随机变化，如图2-1中的“路径损耗和阴影”曲线所示。这条曲线结合了简化的路径损耗模型（2-39）和满足式（2-46）和式（2-50）的对数正态阴影衰落随机过程。按照这个模型，接收功率和发射功率之比的分贝值为

$$\frac{P_r}{P_t}\mathrm{dB}=10\log_{10}K-10\gamma\log_{10}\frac{d}{d_0}-\psi_{dB} \tag{2-51}$$

式中ψ_{dB}是均值为0、方差为$\sigma_{\psi_{dB}}^2$的高斯随机变量。如式（2-51）及图2-1所示，路径损耗随$\log_{10}d$线性下降，下降斜率是10γdB/十倍程，γ为路径损耗指数。阴影衰落变化比较快，和不相关距离X_c具有相同数量级。

51

例2.3和例2.4例证了用表2-3的测量数据建立的路径损耗模型和对数正态阴影衰落模型的混合模型，其中路径损耗符合$K=-31.54$dB、路径损耗指数$\gamma=3.71$的简化路径损耗

模型，阴影衰落服从标准差为$\sigma_{\psi_{dB}}=3.65\text{dB}$均值由路径损耗模型确定的对数正态模型。

2.9 路径损耗和阴影衰落造成的中断率

路径损耗和阴影衰落对无线通信系统的设计有重要意义。无线通信系统一般有一个目标最小接收功率$P_{\min}$，当实际接收功率低于该目标值时，系统性能就会变得不可接受，比如蜂窝系统中电话的音质很差不能听懂。受阴影效应的影响，对于任意给定的到发射机的距离，接收功率是服从对数正态分布的随机值，存在接收功率低于目标值$P_{\min}$的可能性。定义路径损耗和阴影衰落造成的**中断率**（outage probability）$P_{out}(P_{\min},d)$为$P_{out}(P_{\min},d)=p(P_r(d)<P_{\min})$，即距离$d$处的接收功率$P_r(d)$低于目标值$P_{\min}$的概率。对于2.8节所述的路径损耗和阴影的混合模型，有

$$p(P_r(d)\leqslant P_{\min})=1-Q\left(\frac{P_{\min}-(P_t+10\log_{10}K-10\gamma\log_{10}(d/d_0))}{\sigma_{\psi_{dB}}}\right) \tag{2-52}$$

式中的Q函数定义为标准正态随机变量X大于z的概率：

$$Q(z)\triangleq p(X>z)=\int_z^\infty\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-y^2/2}\mathrm{d}y \tag{2-53}$$

Q函数和互补误差函数有如下变换关系：

$$Q(z)=\frac{1}{2}\mathrm{erfc}\left(\frac{z}{\sqrt{2}}\right) \tag{2-54}$$

在不致混淆的情形下，我们将把$P_{out}(P_{\min},d)$简记为P_{out}。

例2.5：请就例2.3和例2.4中路径损耗和阴影的混合模型，求出距离150m处的中断率。假设发射功率为$P_t=10\text{mW}$，要求的最小接收功率为$P_{\min}=-110.5\text{dBm}$。

解：$P_t=10\text{mW}=10\text{dBm}$，因此

$$\begin{aligned}P_{out}(-110.5\text{dBm},150\text{m})&=p(P_r(150\text{m})<-110.5\text{dBm})\\&=1-Q\left(\frac{P_{\min}-(P_t+10\log_{10}K-10\gamma\log_{10}(d/d_0))}{\sigma_{\psi_{dB}}}\right)\\&=1-Q\left(\frac{-110.5-(10-31.54-37.1\log_{10}(150))}{3.65}\right)\\&=0.0121\end{aligned}$$

52

在无线通信系统设计中，1%是一个典型的目标中断率数值。

2.10 小区覆盖范围

蜂窝系统的**小区覆盖范围**（cell coverage area）定义为小区内接收功率超过最小规定值的位置所占的百分比。假设在一个半径为R的圆形小区中有一个基站。为了达到可接受的性能，小区中所有移动台的接收信噪比都必须要达到某个最小规定值。当噪声模型已经确定时，接收信噪比要求转换为最小接收功率$P_{\min}$。基站发射功率可通过小区边界上的平均接收功率$\bar{P}_R$来确定（此处的平均是对阴影衰落平均）。阴影效应将会使小区

中某些位置的接收功率低于$\overline{P}_R$，另一些位置的接收功率高于$\overline{P}_R$。如图2-10所示，图中在基站发射功率固定的条件下，画出了等接收功率线。图中包括了路径损耗加随机阴影和路径损耗加平均阴影两种情形。路径损耗加平均阴影时，对于相同的距离，路径损耗值和阴影衰落平均值是相同的，所以等功率线是以基站为圆心的一个圆。路径损耗加随机阴影时，阴影衰落围绕平均值的随机变化使等功率线像一个变形虫。这样的等功率曲线也暗示了阴影衰落给蜂窝系统设计带来的挑战。我们不可能做到让小区边界上所有的用户都有相同的接收功率。因此，要么提高基站发射功率来使受阴影影响的用户也能达到最小所需的接收功率，但这种做法会增加对相邻小区的干扰；要么有些用户的接收功率就会低于所需的最小值。事实上因为高斯分布有无限长拖尾，所以小区中的任何用户都有一个非零的概率使接收功率达不到最小所需，即便离基站很近也一样。这一点并不奇怪，离基站很近的移动台也可能碰巧就在隧道中，或者正好在某个大型建筑的背后。 53

现在来计算给定路径损耗和阴影衰落条件下的小区覆盖范围，如图2-10所示，在距离基站r处取一个增量区间dA，令$P_r(r)$为考虑路径损耗和阴影之后dA中的接收功率，那么小区内接收功率超过最小所需功率的总面积可通过如下的积分得到：

$$\begin{aligned}C &= \boldsymbol{E}\left[\frac{1}{\pi R^2}\int_{\text{cell area}} 1[P_r(r) > P_{\min}\text{ in d}A]\text{d}A\right] \\ &= \frac{1}{\pi R^2}\int_{\text{cell area}} \boldsymbol{E}[1[P_r(r) > P_{\min}\text{ in d}A]]\text{d}A\end{aligned} \tag{2-55}$$

式中$1[\cdot]$为示性函数。定义P_A为dA处$P_r(r)$大于$P_{\min}$的概率，则$P_A = E[1[\text{d}A\text{处}P_r(r)>P_{\min}]]$。代入式（2-55）并化为极坐标得到

$$C = \frac{1}{\pi R^2}\int_{\text{cell area}} P_A \text{d}A = \frac{1}{\pi R^2}\int_0^{2\pi}\int_0^R P_A r\text{d}r\text{d}\theta \tag{2-56}$$

定义**小区中断率**（outage probability of the cell）为小区中接收功率达不到最小要求值$P_{\min}$的面积百分比：即$P_{\text{out}}^{\text{cell}} = 1 - C$。

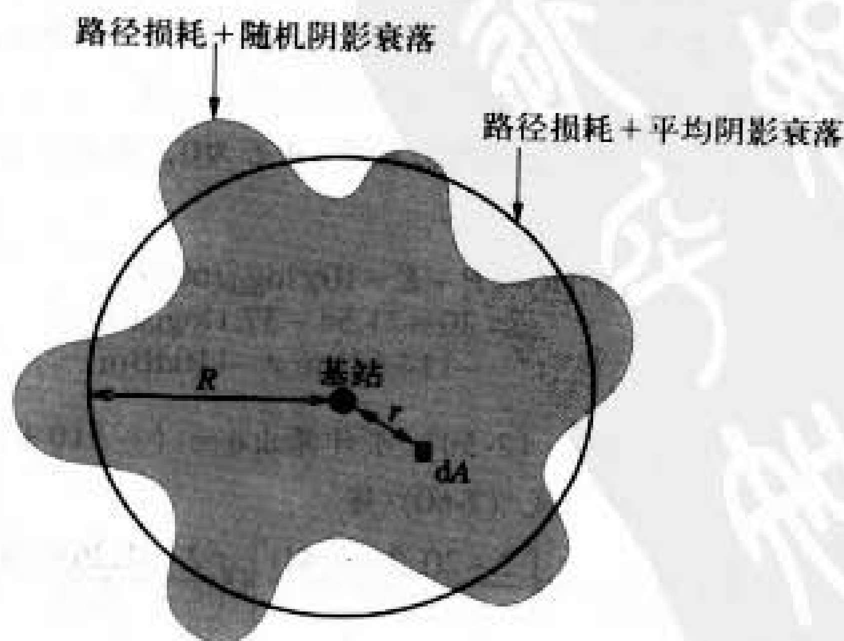


图2-10 等接收功率线

设阴影衰落为对数正态分布，我们得到

$$\begin{aligned} P_A &= p(P_r(r) > P_{\min}) \\ &= Q\left(\frac{P_{\min} - (P_t + 10\log_{10} K - 10\gamma\log_{10}(r/d_0))}{\sigma_{\psi_{dB}}}\right) \\ &= 1 - P_{out}(P_{\min}, r) \end{aligned} \tag{2-57}$$

式中P_{out}是在式（2-52）定义的中断率中代入$d=r$的结果。小区中接收功率低于$P_{\min}$的位置称为中断位置（outage location）。

54

结合式（2-56）和式（2-57）得到[1]

$$C = \frac{2}{R^2}\int_0^R rQ\left(a + b\ln\frac{r}{R}\right)\mathrm{d}r \tag{2-58}$$

其中

$$a = \frac{P_{\min} - \overline{P}_r(R)}{\sigma_{\psi_{dB}}}, \quad b = \frac{10\gamma\log_{10}(\mathrm{e})}{\sigma_{\psi_{dB}}} \tag{2-59}$$

$\overline{P}_r(R) = P_t + 10\log_{10} K - 10\gamma\log_{10}(R/d_0)$是只有路径损耗时，距离基站$R$处的小区边界上的接收功率。积分结果是一个由$a$、$b$组成的闭式解：

$$C = Q(a) + \exp\left[\frac{2-2ab}{b^2}\right]Q\left(\frac{2-ab}{b}\right) \tag{2-60}$$

若目标最小接收功率等于小区边界处的平均接收功率，即若$P_{\min} = \overline{P}_r(R)$，那么$a=0$，覆盖范围简化为

$$C = \frac{1}{2} + \exp\left[\frac{2}{b^2}\right]Q\left(\frac{2}{b}\right) \tag{2-61}$$

简化以后C只和比值$\gamma/\sigma_{\psi dB}$有关。此外由于高斯分布的对称性，在此假设下小区边界处的中断率为$P_{out}(\overline{P}_r(R), R) = 0.5$。

例2.6：某小区的传播模型是例2.3和例2.4中所给出的包括路径损耗与阴影衰落的模型，小区半径600m，基站发射功率$P_t = 100\text{mW} = 20\text{dBm}$。若最小所需接收功率分别为$P_{\min} = -110\text{dBm}$和$P_{\min} = -120\text{dBm}$，请求出相应的小区覆盖范围。

解：先考虑$P_{\min} = -110\text{dBm}$的情形。检验$a$是否为0，以决定是用完整公式（2-60）还是简化公式（2-61）。

$$\begin{aligned} \overline{P}_r(R) &= P_t + K - 10\gamma\log_{10}(600) \\ &= 20 - 31.54 - 37.1\log_{10}(600) \\ &= -114.6\text{dBm} \neq -110\text{dBm} \end{aligned}$$

因此要用式（2-60）。由式（2-59）可计算出$a = (-110+114.6)/3.65 = 1.26$、$b = (37.1\times0.434)/3.65 = 4.41$，代入式（2-60）得

$$C = Q(1.26) + \exp\left[\frac{2-2(1.26\cdot4.41)}{4.41^2}\right]Q\left(\frac{2-(1.26)(4.41)}{4.41}\right) = 0.59$$

1. 注意式（2-57）仅在条件$r \geqslant d_0$下成立，但为了简化分析我们将范围扩展到r所有可能的取值。这个近似对覆盖范围影响不大，因为d_0相对于R一般很小，而且$r<d_0$时的中断率也可以忽略。

对于运营的蜂窝系统来说，这个覆盖值非常低，将存在许多不满意的用户。现在我们再来考虑把最小接收功率的要求降低到$P_{min}=-120$dBm的情形。可算出$a=(-120+114.9)/3.65=-1.479$、$b=4.41$。代入式（2-60）得$C=0.988$，这是一个比较容易接受的覆盖率。

例2.7：设某蜂窝系统的设计中$P_{min}=\overline{P}_r(R)$，即考虑路径损耗和平均阴影衰落时，小区边界处的接收功率等于可接受性能所需要的最小接收功率。请就$\gamma=2, 4, 6$和$\sigma_{\psi dB}=4, 8, 12$的情形分别求出相应的覆盖范围，并说明覆盖范围如何随γ和$\sigma_{\psi dB}$变化。

解：$P_{min}=\overline{P}_r(R)$时$a=0$，覆盖范围可由式（2-61）求出。此时覆盖范围取决于$b=10\gamma\log_{10}(e)/\sigma_{\psi dB}$，而$b$只与$\gamma/\sigma_{\psi dB}$有关。表2-4中列出了不同$\gamma$和$\sigma_{\psi dB}$下，由（2-61）求出的覆盖范围。

55

给定γ时，覆盖范围随$\sigma_{\psi dB}$的减小而增大。这是因为$\sigma_{\psi dB}$较小说明接收信号在平均路径损耗附近的变化较小。当$P_{min}=\overline{P}_r(R)$时，如果没有阴影衰落，覆盖范围是100%，因此若$\sigma_{\psi dB}$减小到0，覆盖范围也将增加到100%。可能有些令人困惑的是，既然γ越大就说明接收信号功率下降得越快，而当给定$\sigma_{\psi dB}$时覆盖范围却随γ的增加而增加。这是因为我们所设定的$P_{min}=\overline{P}_r(R)$已经隐含了功率下降更快这个因素。举例来说，为了保持$P_{min}=\overline{P}_r(R)$，对于$\gamma=6$我们需要比$\gamma=2$更高的发射功率。因此，$\gamma$增加时，为了满足$P_{min}=\overline{P}_r(R)$，发射功率也要增加，从而小区中的接收功率整体提高，覆盖范围提高。

表2-4　不同γ和$\sigma_{\psi dB}$情形下的覆盖范围

γ	$\sigma_{\psi_{dB}}$		
	4	8	12
2	0.77	0.67	0.63
4	0.85	0.77	0.71
6	0.90	0.83	0.77

习题

2.1　对于自由空间路径损耗模型，求使接收功率达到1dBm所需的发射功率。假设载波频率$f=5$GHz、全向天线（$G_l=1$），距离分别为$d=10$m及$d=100$m。

2.2　设两径模型中$h_t=10$m、$h_r=2$m、收发间距$d=100$m，求两路信号的相对延迟。

2.3　对于两径模型，请用泰勒级数近似的方法从式（2-13）得到如下的近似式：

$$\Delta\phi=\frac{2\pi(x+x'-l)}{\lambda}\approx\frac{4\pi h_t h_r}{\lambda d}$$

2.4　请导出两径模型在临界距离d_c之内，接收信号为零的距离值的近似表达式。

2.5　某郊区宏小区的基站位于楼顶，高度$h_t=20$m，移动台高度$h_r=3$m，载频$f_c=2$GHz。求两径模型的临界距离d_c。并请说明这个距离是否适合作为郊区宏小区的半径？

2.6　假设把两径模型中的地面反射换成直射线左侧（或右侧）的建筑物反射，请问这个反射信号的建筑物处在收发端之间的什么位置时，所得模型和原来的两径模型相同？

2.7　已知两径模型下的冲激响应为$h(t)=\alpha_1\delta(t)+\alpha_2\delta(t-0.022)$（时间单位是μs），每个径上的传播符合自由空间模型，地面反射系数是-1，发射机和接收机离地面的高度都是8m，载波频率为900MHz。求α_1、α_2及收发距离。

56

2.8　定向天线可以有效降低干扰、减少多径效应。从本题中我们将会看到，如果天线定向于两径模型中的直射波，就能减弱地面波造成的抵消效应。假设两径模型的参数为$f=900$MHz、$R=-1$、$h_t=50$m、$h_r=2$m、$G_l=1$，请按不同的G_r值1、0.316、0.1和0.01（对应0、−5、−10、−20dB），画出距离范围$d=1$m～100km内分贝接收功率（$10\log_{10}P_r$）和对数距离（$\log_{10}d$）的关系曲线。计算出临界距离$d_c=4h_th_r/\lambda$并标在图中。将图的起点归一化为0dB。并示出折线模型，其接收功率在距离不足h_t时是常值、在距离$h_t<d<d_c$内按$10\log_{10}(d^{-2})$下降、在$d\geqslant d_c$时按$10\log_{10}(d^{-4})$下降。若按功率损耗和对数距离的关系画，就是一个斜率分别是0、2、4的三段折线。注意在距离很大时，很难做到$G_r<<G_l$，因为这将要求天线的方向角度必须非常精细。

2.9　十径模型中，平均功率随距离的下降关系如何？为什么？

2.10　假设十径模型中发射机和接收机高度相同，且都位于宽20m的街道的正中，收发间隔500m。求时延扩展。

2.11　某系统中发射机、接收机和散射体之间的关系如图2-11所示。发射机和接收机的高度$h_t=h_r=4$m，收发相隔距离d。在地平面上，散射体的距离无论横纵都是$0.5d$，即发射机在地平面上的坐标是(0, 0)、接收机是(0, d)、散射体是($0.5d$, $0.5d$)。设雷达截面为20dBm²、$G_s=1$、$f_c=900$MHz。求$d=1$m，10m，100m，1000m时散射信号的路径损耗，并比较该路径损耗与把散射体换成反射系数$R=-1$的反射体时的路径损耗。

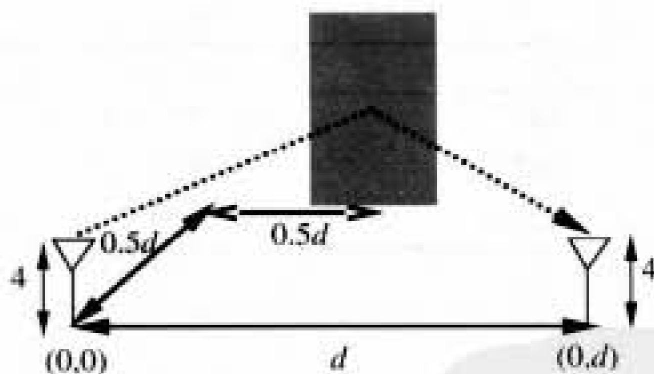


图2-11　题2.11中带散射的系统

2.12　何种条件下，简化路径损耗模型（2-39）和自由空间路径损耗模型（2-7）相同？

2.13　假设接收机的噪声功率在感兴趣的信号带宽范围内是−160dBm。传播模型是参数为$d_0=1$m、$\gamma=4$的简化路径损耗模型，参数K由自由空间模型按全向天线和$f_c=1$GHz确定。若发射功率为$P_t=10$mW，求信噪比为20dB时所对应的最大收发距离。

57

2.14　本题将说明，对于给定的系统设计，采用不同的传播模型会得出不同的信噪比，从而得出不同的链路性能。考虑一个如图2-12所示的可运行在公路沿线的频分带状蜂窝系统，

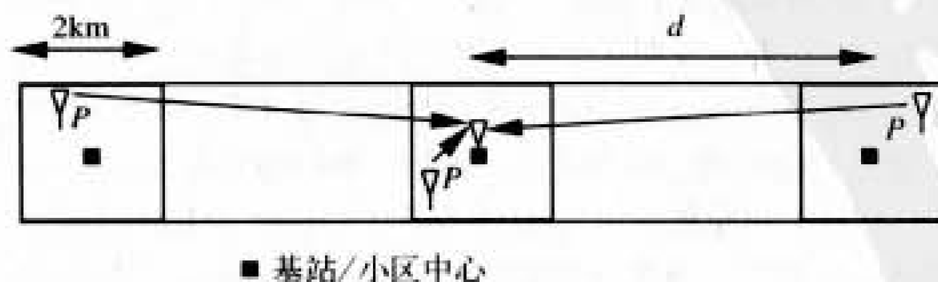


图　2-12

每个小区分配有一定的频率，这些频率又被距离d处的其他小区复用。假设小区形状是边长为2km的正方形，小区中每个移动台的发射功率都是P。请根据下列的传播模型计算能使上行信噪比大于20dB的最小同频小区距离。此处的上行信噪比指移动台发送、基站接收时的最小接收信扰比。假设除邻近的两个同频小区外，更远处同频小区的干扰可以忽略。

（a）信号和干扰都是自由空间模型。

（b）信号和干扰都是简化的路径损耗模型（2-39），模型参数为$d_0=100$m、$K=1$、$\gamma=3$。

（c）信号和干扰都是简化的路径损耗模型，模型参数同为$d_0=100$m、$K=1$，但参数γ对信号是2，对干扰是4。

2.15　假设$f_c=900$MHz、$h_t=20$m、$h_r=5$m、$d=100$m，请分别计算大城市、小城市、郊区和乡村环境下，哈塔模型的路径损耗中值。并定性说明这些路径损耗的差别。

2.16　设两径模型（2-12）中的距离范围是1m~1000m、$h_t=10$m、$h_r=2$m、$f_c=900$MHz、$G_l=G_r=1$。请确定出三段折线近似模型的参数，画出此距离范围内的路径损耗和其折线近似结果。

2.17　在室内衰减模型下，假设发送信号传播了100m，途中经过了衰减分别为15dB、10dB和6dB的三个地板层，还经过了两个双层石膏板隔墙。此外假设参考距离$d_0=1$、路径损耗指数$\gamma=4$、路径损耗常数$K=0$dB。若接收功率是-110dB，求所需的发射功率。

2.18　表2-5列出了一些实测的路径损耗数据。

（a）请用这些数据求出简化路径损耗模型和对数正态阴影衰落的最佳参数。

（b）用所得模型计算2km处的路径损耗。

（c）若只考虑路径损耗时，距离d处的接收功率比正常通信所要求的功率高10dB，求考虑阴影衰落后d处的中断率。

表2-5　题2.18中的路径损耗测量值

离发射机的距离	P_r/P_t	离发射机的距离	P_r/P_t
5m	－60dB	110m	－115dB
25m	－80dB	400m	－135dB
65m	－105dB	1000m	－150dB

2.19　某900MHz蜂窝系统的传播模型是自由空间路径损耗叠加了标准差为$\sigma=6$dB的对数阴影衰落、可接受话音质量要求的信噪比是15dB、基站的发射功率是1W、基站发射天线增益是3dB、移动台天线无增益、移动台在感兴趣的带宽范围内接收到的噪声功率是-40dBm。若要求小区边界上话音质量可接受的几率高于90%，求最大的小区半径。

2.20　本题将以（2-50）的自协方差模型为基础，仿真对数正态衰落随距离变化的过程。仿真方法如课本中所述，先产生一个高斯白过程，再将其通过一个极点在$e^{-\delta/X_c}$的一阶滤波器。请按$X_c=20$m画出仿真得到的衰落过程，假设采样间隔是1m，距离范围是0m~200m。注意对数正态阴影的均值反映路径损耗，故画图时可将结果按均值为0dB进行归一化。

2.21　本题考查阴影衰落的参数对中断率的影响。设某蜂窝系统中的接收功率服从均值为$\mu_{\psi_{dB}}$dBm、标准差为$\sigma_{\psi_{dB}}$dBm的对数正态分布。可接受的性能要求接收功率高于10dBm。

（a）若对数正态分布的$\mu_{\psi_{dB}}=15$dBm、$\sigma_{\psi_{dB}}=8$dBm，问中断率是多少？

(b) 蜂窝系统的典型中断率是1%。若$\sigma_{\psi_{dB}}=4$dBm，求使中断率低于1%所要求的$\mu_{\psi_{dB}}$。

(c) 对于$\sigma_{\psi_{dB}}=12$dBm，重做 (b)。

(d) 当多个基站可以同时接收同一移动台的信号时（如CDMA系统），采用宏分集把多个基站接收到的信号合并起来，就可以降低中断率。请说明其原因。

2.22　请用分部积分法从式（2-59）推导出覆盖范围公式（2-61）。

2.23　某微蜂窝系统的路径损耗符合参数为$\gamma=3$、$d_0=1$m、$K=0$的简化模型，并有$\sigma=4$dB的对数正态阴影。若小区半径为100m、发射功率为80mW、要求的最小接收功率是-100dBm，求小区的覆盖率。

59

2.24　某蜂窝系统的路径损耗符合参数为$\gamma=6$的简化模型，并有$\sigma=8$dB的对数正态阴影。小区边界的接收功率比正常通信最小要求的功率高20dB，求小区覆盖率。

2.25　微蜂窝系统的路径损耗指数典型值在2~6之间，阴影标准差的典型值在4~12之间。若小区边界上的平均接收功率是正常通信最小需要的功率的2倍[1]，求能使覆盖率最大和最小的路径损耗指数和阴影标准差。若路径损耗指数和阴影标准差取在典型值范围的中间，对应的覆盖率是多少？

参考文献

[1] D. Parsons, *The Mobile Radio Propagation Channel,* Wiley, New York, 1994.

[2] M. Pätzold, *Mobile Fading Channels,* Wiley, New York, 2002.

[3] *IEEE J. Sel. Areas Commun.*, Special Issue on Channel and Propagation Modeling for Wireless Systems Design, April 2002 and August 2002.

[4] *IEEE J. Sel. Areas Commun.*, Special Issue on Ultra-Wideband Radio in Multiaccess Wireless Communications, December 2002.

[5] T. S. Rappaport, *Wireless Communications – Principles and Practice,* 2nd ed., Prentice-Hall, Englewood Cliffs, NJ, 2001.

[6] W. C. Jakes, Jr., *Microwave Mobile Communications,* Wiley, New York, 1974 [reprinted by IEEE Press].

[7] A. S. Y. Poon and R. W. Brodersen, "The role of multiple-antenna systems in emerging open access environments," *EE Times Commun. Design Conf.*, October 2003.

[8] T. Kurner, D. J. Cichon, and W. Wiesbeck, "Concepts and results for 3D digital terrain-based wave propagation models: An overview," *IEEE J. Sel. Areas Commun.*, pp. 1002–12, September 1993.

[9] N. Amitay, "Modeling and computer simulation of wave propagation in lineal line-of-sight microcells," *IEEE Trans. Veh. Tech.*, pp. 337–42, November 1992.

[10] J. W. McKown and R. L. Hamilton, Jr., "Ray tracing as a design tool for radio networks," *IEEE Network,* pp. 27–30, November 1991.

[11] K. A. Remley, H. R. Anderson, and A. Weisshar, "Improving the accuracy of ray-tracing techniques for indoor propagation modeling," *IEEE Trans. Veh. Tech.*, pp. 2350–8, November 2000.

[12] H.-J. Li, C.-C. Chen, T.-Y. Liu, and H.-C. Lin, "Applicability of ray-tracing techniques for prediction of outdoor channel characteristics," *IEEE Trans. Veh. Tech.*, pp. 2336–49, November 2000.

[13] A. Domazetovic, L. J. Greenstein, N. Mandayan, and I. Seskar, "A new modeling approach for wireless channels with predictable path geometries," *Proc. IEEE Veh. Tech. Conf.*, September 2002.

1. 作者对该处进行了勘误，这里是根据勘误进行翻译的。——译者注

[14] J. H. Tarng, W.-S. Liu, Y.-F. Huang, and J.-M. Huang, "A novel and efficient hybrid model of radio multipath-fading channels in indoor environments," *IEEE Trans. Ant. Prop.*, pp. 585–94, March 2003.
[15] A. J. Rustako, Jr., N. Amitay, G. J. Owens, and R. S. Roman, "Radio propagation at microwave frequencies for line-of-sight microcellular mobile and personal communications," *IEEE Trans. Veh. Tech.*, pp. 203–10, February 1991.
[16] W. C. Y. Lee, *Mobile Communications Engineering*, McGraw-Hill, New York, 1982.
[17] J.-F. Wagen, "Signal strength measurements at 881 MHz for urban microcells in downtown Tampa," *Proc. IEEE Globecom Conf.*, pp. 1313–17, December 1991.
[18] R. J. C. Bultitude and G. K. Bedal, "Propagation characteristics on microcellular urban mobile radio channels at 910 MHz," *IEEE J. Sel. Areas Commun.*, pp. 31–9, January 1989.
[19] J.-E. Berg, R. Bownds, and F. Lotse, "Path loss and fading models for microcells at 900 MHz," *Proc. IEEE Veh. Tech. Conf.*, pp. 666–71, May 1992.
[20] J. H. Whitteker, "Measurements of path loss at 910 MHz for proposed microcell urban mobile systems," *IEEE Trans. Veh. Tech.*, pp. 125–9, August 1988.
[21] H. Börjeson, C. Bergljung, and L. G. Olsson, "Outdoor microcell measurements at 1700 MHz," *Proc. IEEE Veh. Tech. Conf.*, pp. 927–31, May 1992.
[22] K. Schaubach, N. J. Davis IV, and T. S. Rappaport, "A ray tracing method for predicting path loss and delay spread in microcellular environments," *Proc. IEEE Veh. Tech. Conf.*, pp. 932–5, May 1992.
[23] F. Ikegami, S. Takeuchi, and S. Yoshida, "Theoretical prediction of mean field strength for urban mobile radio," *IEEE Trans. Ant. Prop.*, pp. 299–302, March 1991.
[24] M. C. Lawton and J. P. McGeehan, "The application of GTD and ray launching techniques to channel modeling for cordless radio systems," *Proc. IEEE Veh. Tech. Conf.*, pp. 125–30, May 1992.
[25] J. B. Keller, "Geometrical theory of diffraction," *J. Opt. Soc. Amer.*, 52, pp. 116–30, 1962.
[26] R. J. Luebbers, "Finite conductivity uniform GTD versus knife edge diffraction in prediction of propagation path loss," *IEEE Trans. Ant. Prop.*, pp. 70–6, January 1984.
[27] C. Bergljung and L. G. Olsson, "Rigorous diffraction theory applied to street microcell propagation," *Proc. IEEE Globecom Conf.*, pp. 1292–6, December 1991.
[28] R. G. Kouyoumjian and P. H. Pathak, "A uniform geometrical theory of diffraction for an edge in a perfectly conducting surface," *Proc. IEEE*, pp. 1448–61, November 1974.
[29] G. K. Chan, "Propagation and coverage prediction for cellular radio systems," *IEEE Trans. Veh. Tech.*, pp. 665–70, November 1991.
[30] K. C. Chamberlin and R. J. Luebbers, "An evaluation of Longley–Rice and GTD propagation models," *IEEE Trans. Ant. Prop.*, pp. 1093–8, November 1982.
[31] J. Walfisch and H. L. Bertoni, "A theoretical model of UHF propagation in urban environments," *IEEE Trans. Ant. Prop.*, pp. 1788–96, October 1988.
[32] M. I. Skolnik, *Introduction to Radar Systems*, 2nd ed., McGraw-Hill, New York, 1980.
[33] S. Y. Seidel, T. S. Rappaport, S. Jain, M. L. Lord, and R. Singh, "Path loss, scattering, and multipath delay statistics in four European cities for digital cellular and microcellular radiotelephone," *IEEE Trans. Veh. Tech.*, pp. 721–30, November 1991.
[34] S. T. S. Chia, "1700 MHz urban microcells and their coverage into buildings," *Proc. IEEE Ant. Prop. Conf.*, pp. 504–11, York, U.K., April 1991.
[35] D. Wong and D. C. Cox, "Estimating local mean signal power level in a Rayleigh fading environment," *IEEE Trans. Veh. Tech.*, pp. 956–9, May 1999.
[36] T. Okumura, E. Ohmori, and K. Fukuda, "Field strength and its variability in VHF and UHF land mobile service," *Rev. Elec. Commun. Lab.*, pp. 825–73, September/October 1968.
[37] M. Hata, "Empirical formula for propagation loss in land mobile radio services," *IEEE Trans. Veh. Tech.*, pp. 317–25, August 1980.

60

[38] European Cooperative in the Field of Science and Technical Research EURO-COST 231, "Urban transmission loss models for mobile radio in the 900 and 1800 MHz bands," rev. 2, The Hague, September 1991.

[39] E. McCune and K. Feher, "Closed-form propagation model combining one or more propagation constant segments," *Proc. IEEE Veh. Tech. Conf.*, pp. 1108–12, May 1997.

[40] D. Akerberg, "Properties of a TDMA picocellular office communication system," *Proc. IEEE Globecom Conf.*, pp. 1343–9, December 1988.

[41] V. Erceg, L. J. Greenstein, S. Y. Tjandra, S. R. Parkoff, A. Gupta, B. Kulic, A. A. Julius, and R. Bianchi, "An empirically based path loss model for wireless channels in suburban environments," *IEEE J. Sel. Areas Commun.*, pp. 1205–11, July 1999.

61

[42] M. Feuerstein, K. Blackard, T. Rappaport, S. Seidel, and H. Xia, "Path loss, delay spread, and outage models as functions of antenna height for microcellular system design," *IEEE Trans. Veh. Tech.*, pp. 487–98, August 1994.

[43] P. Harley, "Short distance attenuation measurements at 900 MHz and 1.8 GHz using low antenna heights for microcells," *IEEE J. Sel. Areas Commun.*, pp. 5–11, January 1989.

[44] S. Y. Seidel and T. S. Rappaport, "914 MHz path loss prediction models for indoor wireless communications in multifloored buildings," *IEEE Trans. Ant. Prop.*, pp. 207–17, February 1992.

[45] A. J. Motley and J. M. P. Keenan, "Personal communication radio coverage in buildings at 900 MHz and 1700 MHz," *Elec. Lett.*, pp. 763–4, June 1988.

[46] A. F. Toledo and A. M. D. Turkmani, "Propagation into and within buildings at 900, 1800, and 2300 MHz," *Proc. IEEE Veh. Tech. Conf.*, pp. 633–6, May 1992.

[47] F. C. Owen and C. D. Pudney, "Radio propagation for digital cordless telephones at 1700 MHz and 900 MHz," *Elec. Lett.*, pp. 52–3, September 1988.

[48] S. Y. Seidel, T. S. Rappaport, M. J. Feuerstein, K. L. Blackard, and L. Grindstaff, "The impact of surrounding buildings on propagation for wireless in-building personal communications system design," *Proc. IEEE Veh. Tech. Conf.*, pp. 814–18, May 1992.

[49] C. R. Anderson, T. S. Rappaport, K. Bae, A. Verstak, N. Tamakrishnan, W. Trantor, C. Shaffer, and L. T. Waton, "In-building wideband multipath characteristics at 2.5 and 60 GHz," *Proc. IEEE Veh. Tech. Conf.*, pp. 24–8, September 2002.

[50] L.-S. Poon and H.-S. Wang, "Propagation characteristic measurement and frequency reuse planning in an office building," *Proc. IEEE Veh. Tech. Conf.*, pp. 1807–10, June 1994.

[51] G. Durgin, T. S. Rappaport, and H. Xu, "Partition-based path loss analysis for in-home and residential areas at 5.85 GHz," *Proc. IEEE Globecom Conf.*, pp. 904–9, November 1998.

[52] A. F. Toledo, A. M. D. Turkmani, and J. D. Parsons, "Estimating coverage of radio transmission into and within buildings at 900, 1800, and 2300 MHz," *IEEE Pers. Commun. Mag.*, pp. 40–7, April 1998.

[53] R. Hoppe, G. Wölfle, and F. M. Landstorfer, "Measurement of building penetration loss and propagation models for radio transmission into buildings," *Proc. IEEE Veh. Tech. Conf.*, pp. 2298–2302, April 1999.

[54] E. H. Walker, "Penetration of radio signals into buildings in cellular radio environments," *Bell Systems Tech. J.*, pp. 2719–34, September 1983.

[55] W. C. Y. Lee, *Mobile Communication Design Fundamentals,* Sams, Indianapolis, IN, 1986.

[56] D. M. J. Devasirvathan, R. R. Murray, and D. R. Woiter, "Time delay spread measurements in a wireless local loop test bed," *Proc. IEEE Veh. Tech. Conf.*, pp. 241–5, May 1995.

[57] S. S. Ghassemzadeh, L. J. Greenstein, A. Kavcic, T. Sveinsson, and V. Tarokh, "Indoor path loss model for residential and commercial buildings," *Proc. IEEE Veh. Tech. Conf.*, pp. 3115–19, October 2003.

[58] A. J. Goldsmith, L. J. Greenstein, and G. J. Foschini, "Error statistics of real-time power measurements in cellular channels with multipath and shadowing," *IEEE Trans. Veh. Tech.*, pp. 439–46, August 1994.

[59] A. J. Goldsmith and L. J. Greenstein, "A measurement-based model for predicting coverage areas of urban microcells," *IEEE J. Sel. Areas Commun.*, pp. 1013–23, September 1993.

[60] M. Gudmundson, "Correlation model for shadow fading in mobile radio systems," *Elec. Lett.*, pp. 2145–6, November 7, 1991.

[61] G. L. Stuber, *Principles of Mobile Communications*, 2nd ed., Kluwer, Dordrecht, 2001.

[62] A. Algans, K. I. Pedersen, and P. E. Mogensen, "Experimental analysis of the joint statistical properties of azimuth spread, delay spread, and shadow fading," *IEEE J. Sel. Areas Commun.*, pp. 523–31, April 2002.

[63] M. Marsan and G. C. Hess, "Shadow variability in an urban land mobile radio environment," *Elec. Lett.*, pp. 646–8, May 1990.

62

[64] J. Weitzen and T. Lowe, "Measurement of angular and distance correlation properties of log-normal shadowing at 1900 MHz and its application to design of PCS systems," *IEEE Trans. Veh. Tech.*, pp. 265–73, March 2002.

[65] W. Turin, R. Jana, S. S. Ghassemzadeh, V. W. Rice, and V. Tarokh, "Autoregressive modeling of an indoor UWB channel," *Proc. IEEE Conf. UWB Syst. Technol.*, pp. 71–4, May 2002.

63

第3章　统计多径信道模型

在本章中我们讨论一类能够描述信道中多径相互叠加产生干涉情形的衰落模型。虽然第2章中用来描述确定信道的射线跟踪模型也囊括了多径效应，但在实际当中，确定性模型不足以反映信道的真实特性，必须要采用统计方法来描述多径信道。在本章中，我们用随机时变冲激响应为多径信道建模，以统计方法来描述信道，并讨论其性质。

如果发射一个单脉冲，那么通过多径信道后我们接收到的信号将是一个脉冲序列，序列中的每一个脉冲对应于直射分量或由一个或一簇散射体造成的可分辨多径分量。多径信道的时延扩展将导致接收端信号的明显失真。时延扩展等于最先到达信号分量和最后达到信号分量之间的时间延迟。如果这种时延扩展的大小和信号带宽的倒数值相比很小，那么接收信号在时间上的展宽也比较小。然而当它相对于信号带宽的倒数值相比较大时，接收信号的时域波形就会被明显扩宽，这有可能造成信号严重失真。

多径信道的另一个特征是其时变性。时变性来源于发射机或接收机的运动，这种运动使传送路径中形成多径传播的那些反射点的位置随时间变化。因此如果我们在一个运动中的发射机上不断发射脉冲，就会看到，来自各径的脉冲的幅度、时延乃至各脉冲的多径数目都在不断变化。与由一组固定散射体引起的多径分量相互干涉形成的衰落相比，这种变化在时间上不是很快。本章后面的内容首先给出一个通用的时变信道冲激响应模型，无论是快变或慢变信道，这都是一个适用的模型。然后将这个模型用到窄带衰落中。所谓窄带衰落是指衰落信道的带宽小于时延扩展的倒数。对于窄带模型，我们将假设传播环境是准静态的，传播环境中的多径数目固定，每径的路径损耗和阴影衰落也都固定。然后我们研究这种环境下短距离内的信号变化，即小尺度变化，这种变化的原因是多径分量的相互干涉。我们还将用对时变冲激响应进行二维变换的方法来获得宽带多径信道的统计特性。本章最后讨论了离散时间信道模型和空时信道模型。

3.1　时变信道的冲激响应

和第2章中一样，假设发送的信号为：

$$s(t)=\mathrm{Re}\{u(t)\mathrm{e}^{\mathrm{j}2\pi f_c t}\}=\mathrm{Re}\{u(t)\}\cos(2\pi f_c t)-\mathrm{Im}\{u(t)\}\sin(2\pi f_c t) \tag{3-1}$$

其中$u(t)$是$s(t)$的等效基带信号，其带宽为B_u，f_c是载波频率。忽略噪声，接收信号是直射信号分量及所有可分辨多径分量之和：

$$r(t)=\mathrm{Re}\left\{\sum_{n=0}^{N(t)}\alpha_n(t)u(t-\tau_n(t))\mathrm{e}^{\mathrm{j}(2\pi f_c(t-\tau_n(t))+\phi_{D_n}(t))}\right\} \tag{3-2}$$

其中$n=0$对应直射路径。此表达式中的未知量包括：可分辨多径的数目$N(t)$（关于$N(t)$，我们在下文中会详细探讨）、各径的路径长度$r_n(t)$、相应的时延$\tau_n(t)=r_n(t)/c$、多普勒相移$\phi_{D_n}(t)$、幅度$\alpha_n(t)$。

每一个可分辨径可能是由单个反射体形成的，也可能是由一簇时延基本相同的反射体形成的，如图3-1所示。如果第n径仅由单一反射体形成，那么它的幅度$\alpha_n(t)$由该路径

的路径损耗和阴影衰落确定，时延$\tau_n(t)$引起的相移为$e^{-j2\pi f_c \tau_n(t)}$，多普勒频移为$f_{D_n}(t)=v\cos\theta_n(t)/\lambda$，其中$\theta_n(t)$是信号的到达方向和接收机移动方向之间的夹角。多普勒频移导致的多普勒相移为$\phi_{D_n}(t)=\int_t 2\pi f_{D_n}(\tau)\mathrm{d}\tau$。第$n$径也有可能是由反射体簇形成的[1]。对于两个时延分别是τ_1、τ_2的径，若其时延之差远大于信号带宽的倒数，即$|\tau_1-\tau_2|>>B_u^{-1}$，则称这两个径是可分辨的。当多径分量不满足这个条件时，$\mu(t-\tau_1)\approx\mu(t-\tau_2)$，因此在接收端不可能把它们分离出来。几个不可分辨信号分量合成为一个时延$\tau\approx\tau_1\approx\tau_2$的单径分量。这些分量相互干涉，使合成信号的幅度发生快速变化。宽带信道一般都有可分辨的多径分量，式（3-2）中的每项对应于一个单反射体形成的分量或者是多个不可分辨分量的和。窄带信道下，式（3-2）中的每一项都是不可分辨的。

65

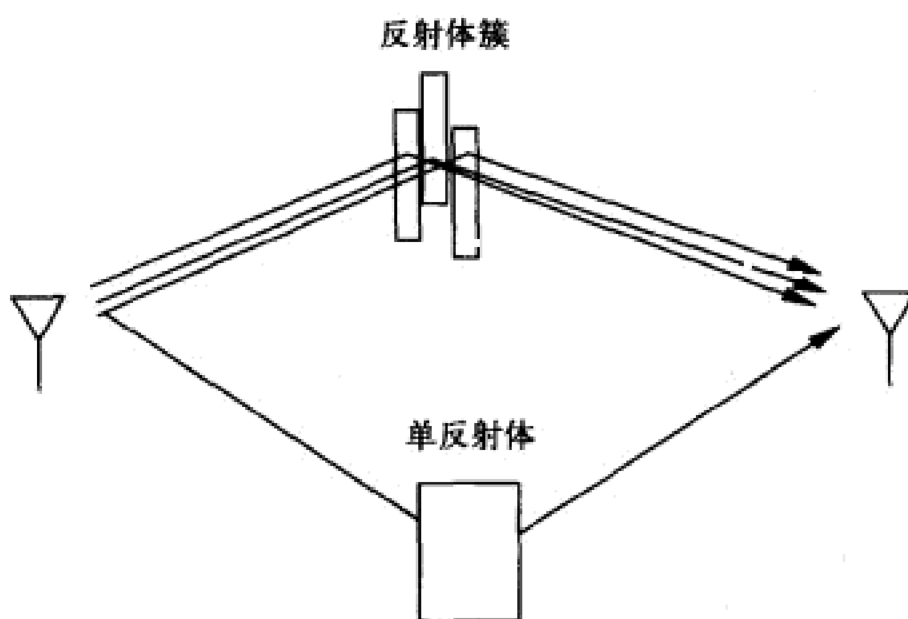


图3-1 单反射体和反射体簇

对于每个可分辨径，其参数$\alpha_n(t)$、$\tau_n(t)$和$\phi_{D_n}(t)$都是随时间变化的，我们将其描述为平稳遍历随机过程。这样，接收信号也是平稳遍历随机过程。对于宽带信道来说，式（3-2）中的每一项对应于一个单反射体，这些参数将随着传播环境的变化而缓慢变化。对于窄带信道来说，式（3-2）中的每一项都是多个不可分辨径的和，多径分量的相互干涉将使这些参数变化很快，大约在信号波长的距离上变化。

令

$$\phi_n(t)=2\pi f_c\tau_n(t)-\phi_{D_n}(t) \tag{3-3}$$

那么接收信号$r(t)$可以简化为

$$r(t)=\mathrm{Re}\left\{\left[\sum_{n=0}^{N(t)}\alpha_n(t)\mathrm{e}^{-\mathrm{j}\phi_n(t)}u(t-\tau_n(t))\right]\mathrm{e}^{\mathrm{j}2\pi f_c t}\right\} \tag{3-4}$$

由于$\alpha_n(t)$取决于路径损耗和阴影衰落，而$\phi_n(t)$取决于时延和多普勒频移，所以一般可假设这两个随机过程是相互独立的。

等效基带输入信号$u(t)$和时变信道的等效基带冲激响应$c(\tau,t)$卷积，再上变频到载波

1. 表面粗糙的单个反射体也能产生出时延差别很小的多个径。

频率就得到接收信号$r(t)$：[1]

66

$$r(t)=\mathrm{Re}\left\{\left(\int_{-\infty}^{\infty}c(\tau,t)u(t-\tau)\,\mathrm{d}\tau\right)\mathrm{e}^{\mathrm{j}2\pi f_c t}\right\} \tag{3-5}$$

要注意到$c(\tau,t)$有两个时间参数：t是在接收端观察到脉冲响应的时刻，$(t-\tau)$是向信道发射冲激脉冲的时刻。如果在时刻t，信道中不存在时延为$\tau_n(t)=\tau$的反射体，那么$c(\tau,t)=0$。时变信道冲激响应的这种定义初看似乎不太符合直觉，但只有这样定义才能和作为特例情形的时不变信道一致。对于时不变信道，$c(\tau,t)=c(\tau,t+T)$，就是说，$(t-\tau)$时刻发送的冲激在t时刻的响应等效于$(t+T-\tau)$时刻发送的冲激在$(t+T)$时刻的响应。令$T=-t$，则$c(\tau,t)=c(\tau,t-t)=c(\tau)$，这个$c(\tau)$就是标准的时不变信道的冲激响应，即0时刻发送的冲激在τ时刻的响应[2]。

由式（3-4）和式（3-5）可见，在$(t-\tau)$时刻发送的冲激在t时刻的等效基带响应$c(\tau,t)$必然由下式决定：

$$c(\tau,t)=\sum_{n=0}^{N(t)}\alpha_n(t)\mathrm{e}^{-\mathrm{j}\phi_n(t)}\delta(\tau-\tau_n(t)) \tag{3-6}$$

将式（3-6）代入式（3-5）得到式（3-4），就能验证出式（3-6）就是等效基带信道的时变冲激响应：

$$\begin{aligned}r(t)&=\mathrm{Re}\left\{\left[\int_{-\infty}^{\infty}c(\tau,t)u(t-\tau)\,\mathrm{d}\tau\right]\mathrm{e}^{\mathrm{j}2\pi f_c t}\right\}\\&=\mathrm{Re}\left\{\left[\int_{-\infty}^{\infty}\sum_{n=0}^{N(t)}\alpha_n(t)\mathrm{e}^{-\mathrm{j}\phi_n(t)}\delta(\tau-\tau_n(t))u(t-\tau)\,\mathrm{d}\tau\right]\mathrm{e}^{\mathrm{j}2\pi f_c t}\right\}\\&=\mathrm{Re}\left\{\left[\sum_{n=0}^{N(t)}\alpha_n(t)\mathrm{e}^{-\mathrm{j}\phi_n(t)}\left(\int_{-\infty}^{\infty}\delta(\tau-\tau_n(t))u(t-\tau)\,\mathrm{d}\tau\right)\right]\mathrm{e}^{\mathrm{j}2\pi f_c t}\right\}\\&=\mathrm{Re}\left\{\left[\sum_{n=0}^{N(t)}\alpha_n(t)\mathrm{e}^{-\mathrm{j}\phi_n(t)}u(t-\tau_n(t))\right]\mathrm{e}^{\mathrm{j}2\pi f_c t}\right\}\end{aligned}$$

式中最后一个等式使用了冲激函数的抽样性质：

$$\int\delta(\tau-\tau_n(t))u(t-\tau)\,\mathrm{d}\tau=\delta(t-\tau_n(t))*u(t)=u(t-\tau_n(t))$$

有些信道模型假设多径时延是连续的，这样式（3-6）中的求和就变为了积分，并可简化为对应不同多径时延τ的时变复包络：

67

$$c(\tau,t)=\int\alpha(\xi,t)\mathrm{e}^{-\mathrm{j}\phi(\xi,t)}\delta(\tau-\xi)\,\mathrm{d}\xi=\alpha(\tau,t)\mathrm{e}^{-\mathrm{j}\phi(\tau,t)} \tag{3-7}$$

1. 带通信号和带通系统的等效基带表示见附录A。
2. 按定义，$c(\tau,0)$是$-\tau$时刻发送的冲激在0时刻的响应。因为信道是时不变的，所以它等于0时刻发送的冲激在τ时刻的响应。

为了深化对时变冲激响应的理解，我们来看图3-2所示的例子，其中每个径对应于一个单反射体。在时刻t_1，接收信号由三个径组成，其幅度、相位和时延是$(\alpha_i, \phi_i, \tau_i)$，$i=1, 2, 3$。$t_1$时刻能同时收到$(t_1-\tau_i)$ $(i=1, 2, 3)$时刻分别发送的冲激，而收不到其他任何时刻发送的冲激（因为没有相应时延的路径）。t_1时刻的时变冲激响应为

$$c(\tau, t_1) = \sum_{n=0}^{2} \alpha_n e^{-j\phi_n} \delta(\tau - \tau_n) \tag{3-8}$$

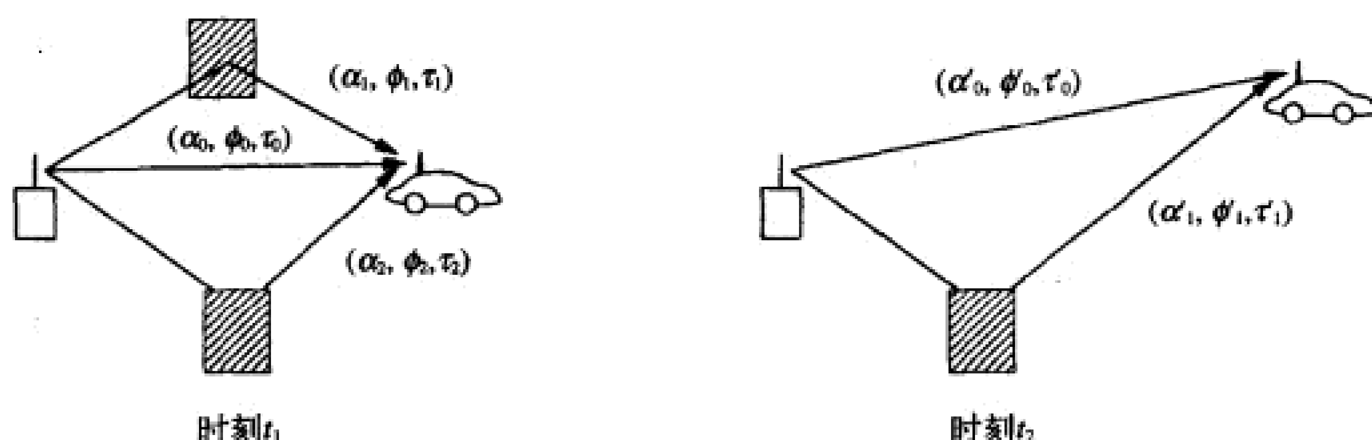


图3-2 不同观测时刻的多径情形

图3-3给出了$t=t_1$时刻的信道冲激响应。图3-2中t_2时刻接收信号包括两个径，其幅度、相位和时延分别为$(\alpha'_i, \phi'_i, \tau'_i)$，$i=1, 2$。这样，$t_2$时刻能收到在$t_2-\tau'_1$和$t_2-\tau'_2$时刻发送的冲激，不能收到其他任何时刻发送冲激，同样是因为没有相应时延的路径。t_2时刻的时变冲激响应示于图3-3，其表达式为：

$$c(\tau, t_2) = \sum_{n=0}^{1} \alpha'_n e^{-j\phi'_n} \delta(\tau - \tau'_n) \tag{3-9}$$

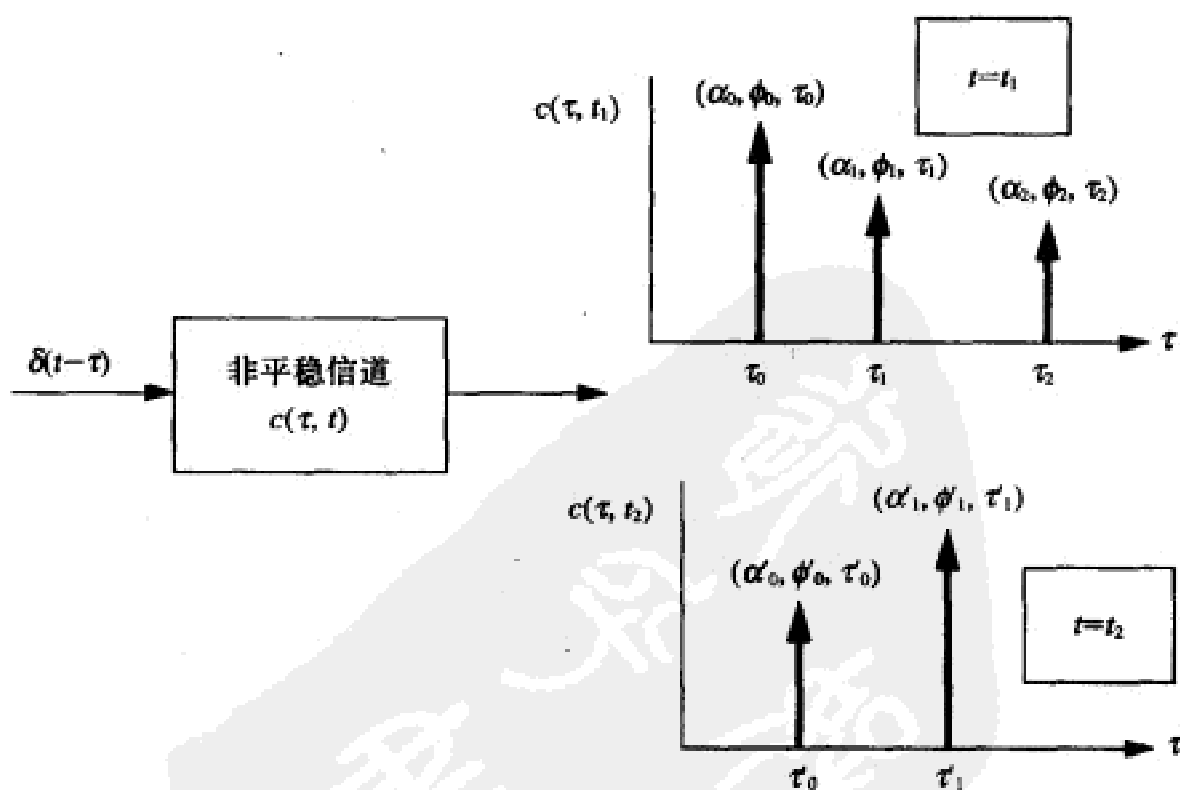


图3-3 非平稳信道的冲激响应

对于时不变信道，$c(\tau, t)$中的t为常量，$c(\tau, t)=c(\tau)$只是τ的函数。对于离散多径信道，$c(\tau)$可表示为：

68

$$c(\tau)=\sum_{n=0}^{N}\alpha_n e^{-j\phi_n}\delta(\tau-\tau_n) \tag{3-10}$$

对于连续多径信道，$c(\tau)=\alpha(\tau)e^{-j\phi(\tau)}$。在静态信道中，$t_1$时刻发送的冲激产生的响应就是$t_2\neq t_1$时刻发送的冲激产生的响应的时间平移。

例3.1：某工厂中有一个靠近传送带的无线局域网。发射机和接收机之间存在一个增益为α_0、相移为ϕ_0、时延为τ_0的直射路径。每隔T_0秒，有一个金属物件从传送带传下来，它将产生一个增益为α_1、相移为ϕ_1、时延为τ_1的额外反射路径。请给出信道的时变冲激响应$c(\tau,t)$。

解： *当$t\neq nT_0(n=1, 2,\cdots)$时，信道冲激响应只包括直射路径的冲激响应。当$t=nT_0$时，信道冲激响应包括直射路径和反射路径的冲激响应。因此，$c(\tau,t)$可以表示为*

$$c(\tau,t)=\begin{cases}\alpha_0 e^{j\phi_0}\delta(\tau-\tau_0) & t\neq nT_0\\ \alpha_0 e^{j\phi_0}\delta(\tau-\tau_0)+\alpha_1 e^{j\phi_1}\delta(\tau-\tau_1) & t=nT_0\end{cases}$$

注意，对于典型的载波频率来说，第n径会满足$f_c\tau_n(t)>>1$。例如室内系统的典型时延是$\tau_n=50$ns，若$f_c=1$GHz，则$f_c\tau_n=50>>1$。室外系统的多径时延要比50ns大得多，因而也同样满足这一点。当$f_c\tau_n(t)>>1$时，第n条径上路径时延的微小变化将导致相位$\phi_n(\tau)=2\pi f_c\tau_n(t)-\phi_{D_n}-\phi_0$极大的变化。各径上相位的快速变化将造成剧烈的干涉现象，从而使接收信号强度发生快速的变化。这种现象叫作*衰落*（fading），下面几节将对此进行详细的讨论。

69

多径传播对接收信号的影响取决于直射径与多径的时延扩展相对于信号带宽倒数的大小。如果信道时延扩展相对很小，那么直射分量和多径分量一般是不可分辨的，这就是下一节将要介绍的窄带衰落模型。如果时延扩展相对比较大，那么直射分量和其他多径分量一般可以分解为若干个可分辨的分量，这就是3.3节将要讨论的宽带衰落模型。注意宽带衰落中的某些可分辨分量是由一些不可分辨的分量组成的。一般相对与解调器同步的那个分量来测量时延扩展。这样，对于时不变信道模型（3-10），如果解调器同步于具有最小时延τ_0的直射分量，那么时延扩展为常数$T_m=\max_n[\tau_n-\tau_0]$。如果解调器同步于时延等于平均时延$\bar{\tau}$的多径分量，则时延扩展是$T_m=\max_n[\tau_n-\bar{\tau}]$。时变信道中的多径时延是随时间随机变化的，因而时延扩展T_m是一个随机变量。此外，某些径上的接收功率可能显著低于其他径，计算时延扩展时，是否应该考虑这些径就成为一个问题。特别地，当某个径的功率低于背景噪声时，它对时延扩展的影响应该很小。为了解决这些问题，一般用3.3.1节所定义的功率时延谱来确定时延扩展。时延扩展的两个特征值：平均时延扩展和均方根时延扩展，都是由功率时延谱确定的。有时还会用到时延扩展的其他特征，如超额时延扩展、时延窗口、时延间隔等，这方面的内容请参见文献[1]的5.4.1节及文献[2]的6.7.1节。精确地定义时延扩展对于理解其在多径信道中的影响并不重要，只要它大致体现了主要的多径成分的时延即可。下文中主要用均方根时延扩展，也可以换用其他合理的定义。当解调器同步于具有平均时延的分量时，均方根时延扩展

是最常用的定义，因为均方根值是对围绕平均值的变化程度最好的测度。信道的时延扩展与传播环境密切相关，时延扩展的典型值范围：在室内环境中是10ns~1000ns，在郊区环境下是200ns~2000ns，在城区环境下是1μs~30μs（见文献[1]的第5章）。

3.2 窄带衰落模型

假定信道的时延扩展T_m远远小于发送信号带宽B的倒数，即$T_m \ll B^{-1}$。前面已经指出，时变信道的时延扩展T_m一般定义为均方根时延扩展，但也可以采用其他定义。不管是什么定义，满足$T_m \ll B^{-1}$都意味着所有径的时延τ_i都小于T_m，因此所有径都近似有$u(t-\tau_i)\approx u(t)$，这样式（3-4）可改写为

$$r(t)=\mathrm{Re}\left\{u(t)\mathrm{e}^{\mathrm{j}2\pi f_c t}\left(\sum_{n=0}^{N(t)}\alpha_n(t)\mathrm{e}^{-\mathrm{j}\phi_n(t)}\right)\right\} \tag{3-11}$$

此式表明，接收信号和发送信号在表达形式上只相差大括号中的那个复系数。在$T_m \ll B^{-1}$的窄带条件下，这个复系数和发送信号$s(t)$及其等效基带信号$u(t)$无关。为了更好地描述多径造成的这个随机复系数的特性，我们假设$s(t)$是相移为ϕ_0的未调制载波：

$$s(t)=\mathrm{Re}\{\mathrm{e}^{\mathrm{j}(2\pi f_c t+\phi_0)}\}=\cos(2\pi f_c t+\phi_0) \tag{3-12}$$

$s(t)$对于任意时延扩展T_m都是窄带的。此时，接收信号为

$$r(t)=\mathrm{Re}\left\{\left[\sum_{n=0}^{N(t)}\alpha_n(t)\mathrm{e}^{-\mathrm{j}\phi_n(t)}\right]\mathrm{e}^{\mathrm{j}2\pi f_c t}\right\}=r_I(t)\cos 2\pi f_c t-r_Q(t)\sin 2\pi f_c t \tag{3-13}$$

70

式中的同相分量和正交分量的表达式为

$$r_I(t)=\sum_{n=0}^{N(t)}\alpha_n(t)\cos\phi_n(t) \tag{3-14}$$

$$r_Q(t)=\sum_{n=0}^{N(t)}\alpha_n(t)\sin\phi_n(t) \tag{3-15}$$

其中相位项受时延、多普勒以及发送相位ϕ_0的共同影响：

$$\phi_n(t)=2\pi f_c\tau_n(t)-\phi_{D_n}-\phi_0 \tag{3-16}$$

当$N(t)$很大时，由中心极限定理及$\alpha_n(t)$和$\phi_n(t)$相互独立这一特点可知，$r_I(t)$和$r_Q(t)$近似于联合高斯随机过程。若$\alpha_n(t)$为瑞利分布、$\phi_n(t)$在$[-\pi,\pi]$内均匀分布，那么对于比较小的N，同样可认为$r_I(t)$和$r_Q(t)$是联合高斯随机过程。这个结论也可由中心极限定理得到，此时，第n径分量包含来自同一反射体簇的大量不可分辨径[3]。

3.2.1 自相关、互相关及功率谱密度

下面推导接收信号的同相分量$r_I(t)$和正交分量$r_Q(t)$的自相关函数和互相关函数。推导中使用的传播模型假设没有直射路径，因此所得结果不一定适用于直射路径起主要作用的情形。本节始终遵循下列假定：幅度$\alpha_n(t)$、多径时延$\tau_n(t)$和多普勒频移$f_{D_n}(t)$在我们关注的时间范围内变化足够慢，以致可以看成是常数：$\alpha_n(t)\approx\alpha_n$、$\tau_n(t)=\tau_n$、$f_{D_n}(t)=f_{D_n}$。

这一假设在每一条可分辨径都来自单反射体的情况下是成立的。在上述假设下，多普勒相移[1]为$\phi_{D_n}(t)=\int_t 2\pi f_{D_n}dt=2\pi f_{D_n}t$，第$n$径的相位为$\phi_n(t)=2\pi f_c\tau_n-2\pi f_{D_n}t-\phi_0$。

现在我们做一个重要假设：第n径的相位$\phi_n(t)$中，$2\pi f_c\tau_n$这一项比其他项变化要快得多。这是一个合理的假设，因为f_c很大，所以多径时延τ_n的微小变化就可以导致$2\pi f_c\tau_n$发生360°的变化。从这个假设可以得出$\phi_n(t)$在$[-\pi,\pi]$内均匀分布。这样

71

$$E[r_I(t)]=E\left[\sum_n\alpha_n\cos\phi_n(t)\right]=\sum_n E[\alpha_n]E[\cos\phi_n(t)]=0 \tag{3-17}$$

式中第二个等式是由于α_n和ϕ_n相互独立，第三个等式是由于ϕ_n均匀分布。同理可得$E[r_Q(t)]=0$。这样，对于接收信号有$E[r(t)]=0$，说明它是一个零均值高斯过程。如果存在强直射路径，那么接收信号的相位主要受直射分量相位的影响。接收端可以确定出直射信号的相位，因此总相位不再满足均匀分布的假设。

再来看同相分量和正交分量的自相关函数和互相关函数。由于α_n和ϕ_n独立，ϕ_n和ϕ_m（$n\neq m$）独立，并且ϕ_n为均匀分布，故有

$$\begin{aligned}E[r_I(t)r_Q(t)]&=E\left[\sum_n\alpha_n\cos\phi_n(t)\sum_m\alpha_m\sin\phi_m(t)\right]\\&=\sum_n\sum_m E[\alpha_n\alpha_m]E[\cos\phi_n(t)\sin\phi_m(t)]\\&=\sum_n E[\alpha_n^2]E[\cos\phi_n(t)\sin\phi_n(t)]\\&=0\end{aligned} \tag{3-18}$$

即$r_I(t)$和$r_Q(t)$不相关。又因为它们是联合高斯过程，所以它们也是相互独立的。

可用类似推导得到$r_I(t)$的自相关函数为

$$A_{r_I}(t,t+\tau)=E[r_I(t)r_I(t+\tau)]=\sum_n E[\alpha_n^2]E[\cos\phi_n(t)\cos\phi_n(t+\tau)] \tag{3-19}$$

代入$\phi_n(t)=2\pi f_c\tau_n-2\pi f_{D_n}t-\phi_0$和$\phi_n(t+\tau)=2\pi f_c\tau_n-2\pi f_{D_n}(t+\tau)-\phi_0$，得到

$$\begin{aligned}&E[\cos\phi_n(t)\cos\phi_n(t+\tau)]\\&\quad=0.5E[\cos 2\pi f_{D_n}\tau]+0.5E[\cos(4\pi f_c\tau_n-4\pi f_{D_n}t-2\pi f_{D_n}\tau-2\phi_0)]\end{aligned} \tag{3-20}$$

由于$4\pi f_c\tau_n$相对于其他项变化很快且为均匀分布，故式（3-20）中第二项的数学期望为零。再考虑到已假设$f_{D_n}=(v\cos\theta_n)/\lambda$是固定的，这样[2]

$$\begin{aligned}A_{r_I}(t,t+\tau)&=\frac{1}{2}\sum_n E[\alpha_n^2]E[\cos(2\pi f_{D_n}\tau)]\\&=\frac{1}{2}\sum_n E[\alpha_n^2]E\left[\cos\left(\frac{2\pi v\tau}{\lambda}\cos\theta_n\right)\right]\end{aligned} \tag{3-21}$$

可以发现，自相关函数$A_{r_I}(t,t+\tau)$只依赖于τ，即$A_{r_I}(t,t+\tau)=A_{r_I}(\tau)$，因此$r_I(t)$是广义平稳随机过程。

---

1. 为简单起见，我们假设$t=0$时的多普勒相位是零。这一假设不影响分析。
2. 原书公式有误，此处已更正。——译者注

同理可证，正交分量也是广义平稳过程，其自相关函数为$A_{r_Q}(\tau)=A_{r_I}(\tau)$。此外，同相分量和正交分量的互相关函数只和时间差$\tau$有关，其表达式为[1]

$$\begin{aligned}A_{r_I,r_Q}(t,t+\tau)&=A_{r_I,r_Q}(\tau)=E\left[r_I(t)r_Q(t+\tau)\right]\\&=\frac{1}{2}\sum_n E[\alpha_n^2]E\left[\sin\left(\frac{2\pi v\tau}{\lambda}\cos\theta_n\right)\right]=-E[r_Q(t)r_I(t+\tau)]\end{aligned} \tag{3-22}$$

72

利用上述结果可以证明，接收信号$r(t)=r_I(t)\cos(2\pi f_c t)-r_Q(t)\sin(2\pi f_c t)$也是广义平稳过程，其自相关函数为

$$A_r(\tau)=E[r(t)r(t+\tau)]=A_{r_I}(\tau)\cos(2\pi f_c\tau)+A_{r_I,r_Q}(\tau)\sin(2\pi f_c\tau) \tag{3-23}$$

通过对传播环境增加假设条件可进一步简化式（3-21）和式（3-22）。我们主要考虑由Clarke[4]提出，并由Jakes[5]进一步发展的*均匀散射环境*（uniform scattering environment）。在这个模型中，散射体密集分布在各个角度上，如图3-4所示。假设有N个多径分量，第n径的到达角为$\theta_n=n\Delta\theta$，$\Delta\theta=2\pi/N$。同时假定各径有相同的接收功率，即$E[\alpha^2(n)]=2P_r/N$，P_r为总接收功率。这样，式（3-21）可写成[2]

$$A_{r_I}(\tau)=\frac{P_r}{N}\sum_{n=1}^{N}\cos\left(\frac{2\pi v\tau}{\lambda}\cos n\Delta\theta\right) \tag{3-24}$$

代入$N=2\pi/\Delta\theta$得到

$$A_{r_I}(\tau)=\frac{P_r}{2\pi}\sum_{n=1}^{N}\cos\left(2\pi v\tau\cos\frac{n\Delta\theta}{\lambda}\right)\Delta\theta \tag{3-25}$$

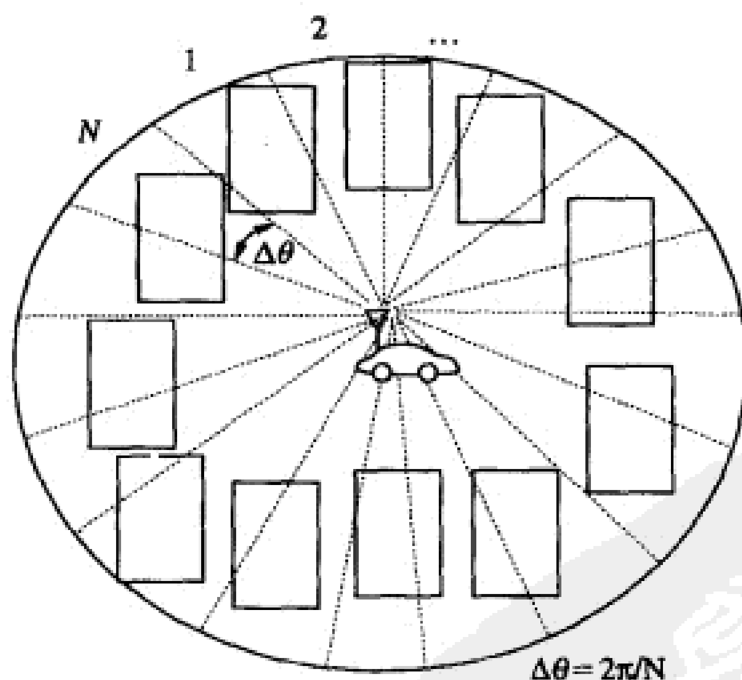


图3-4 密集散射环境

当散射体的个数趋于无穷大的时候，所有方向上的散射都是均匀的。将$N\to\infty$，$\Delta\theta\to 0$代

1. 原书公式有误，此处已更正。——译者注
2. 原书公式有误，此处已更正。——译者注

入式（3-25），求和运算就变成了积分运算：

73

$$A_{r_I}(\tau)=\frac{P_r}{2\pi}\int_0^{2\pi}\cos\left(2\pi v\tau\cos\frac{\theta}{\lambda}\right)\mathrm{d}\theta=P_rJ_0(2\pi f_D\tau)$$
$$J_0(x)=\frac{1}{\pi}\int_0^{\pi}\mathrm{e}^{-\mathrm{j}x\cos\theta}\,\mathrm{d}\theta \tag{3-26}$$

其中 $J_0(x)=\frac{1}{\pi}\int_0^{\pi}\mathrm{e}^{-\mathrm{j}x\cos\theta}\mathrm{d}\theta$ 是零阶贝塞尔函数[1]。对于均匀散射环境，还可以类似得到

$$A_{r_I,r_Q}(\tau)=\frac{P_r}{2\pi}\int\sin\left(2\pi v\tau\cos\frac{\theta}{\lambda}\right)\mathrm{d}\theta=0 \tag{3-27}$$

图3-5示出了$J_0(2\pi f_D\tau)$ 的曲线。图中有这样几点值得注意：首先可以注意到$f_D\tau\approx0.4$时，也即$v\tau=0.4\lambda$时，自相关值为0。就是说，在θ_n均匀分布的假设下，信号在相隔大约半波长的距离后变得不相关。许多实际系统经常用这个近似来确定参数。例如，在第7章中我们将会看到，利用几个天线上独立的衰落路径能够消除衰落的一些不良影响。为此，天线的间隔必须能保证每个接收天线上的衰落彼此独立。按照上述分析，天线间隔就应该是0.4λ。图3-5中另一个值得注意的地方是，经过第一次不相关后，曲线呈现出相关，再经过一定距离后又变得不相关。就是说，我们不能假设间隔距离经过0.4λ后，信号会保持不相关。这个现象说明马尔科夫模型对于瑞利衰落不完全准确。不过许多系统中的分析表明，相关值低于0.5时的性能和不相关相比差不了多少（见文献[6]的9.6.5节）。此时可近似认为一旦信号相关值接近0之后（如间距超过半个波长），将继续保持不相关，从而可将瑞利过程近似为马尔科夫模型。

74

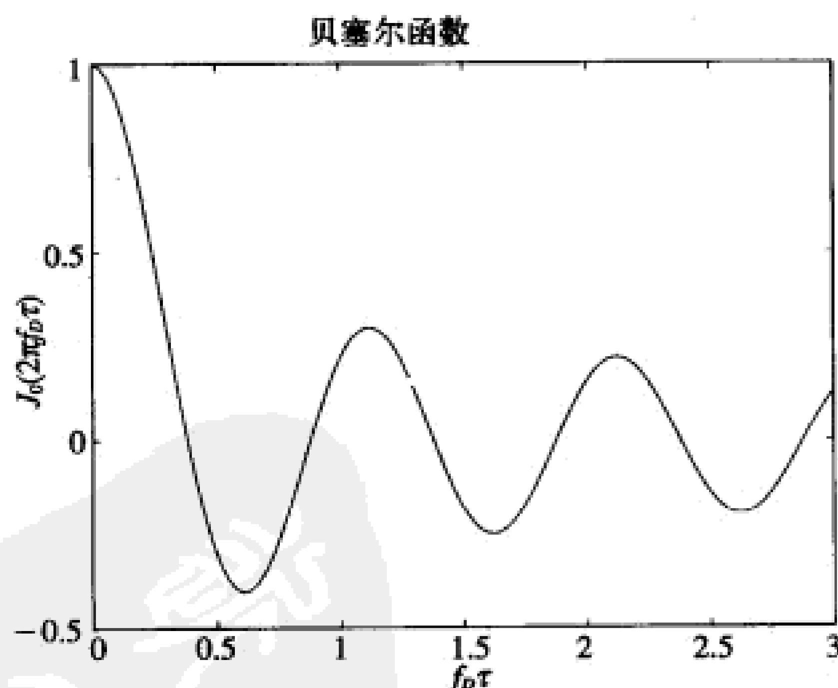


图3-5　以$f_D\tau$为自变量的贝塞尔函数

1. 式（3-26）还有一种等价的导出方式：令式（3-21）中的θ_n为均匀分布的随机变量，再对θ_n求数学期望也得到式（3-26）。从物理模型看，只有在密集散射的环境中，θ_n才是均匀分布的。

分别对自相关函数中的参数τ做傅里叶变换可得到$r_I(t)$和$r_Q(t)$的功率谱密度$S_{r_I}(f)$和$S_{r_Q}(f)$。因为自相关函数相同，所以功率谱密度也相同，为

$$S_{r_I}(f)=S_{r_Q}(f)=\mathcal{F}[A_{r_I}(\tau)]=\begin{cases}\dfrac{2P_r}{\pi f_D}\dfrac{1}{\sqrt{1-(f/f_D)^2}} & |f|\leqslant f_D\\ 0 & \text{其他}\end{cases} \tag{3-28}$$

图3-6示出了功率谱密度曲线。

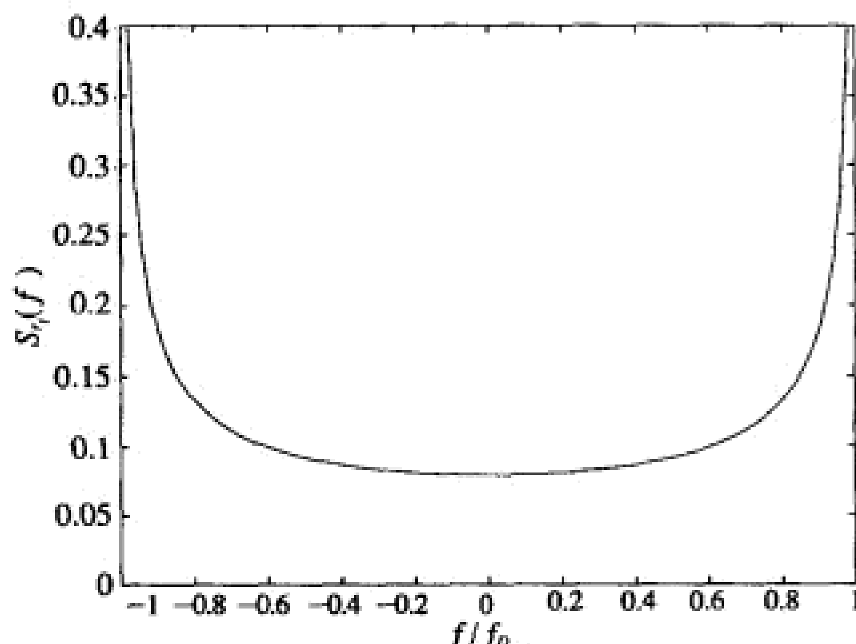


图3-6 同相分量和正交分量的功率谱密度曲线：$S_{r_I}(f)=S_{r_Q}(f)$

利用式（3-23）、$A_{r_I,r_Q}(\tau)=0$及式（3-28），通过傅里叶变换的简单性质可得到均匀散射情况下接收信号的功率谱密度为

$$\begin{aligned}S_r(f)&=\mathcal{F}[A_r(\tau)]=0.25[S_{r_I}(f-f_c)+S_{r_I}(f+f_c)]\\&=\begin{cases}\dfrac{P_r}{2\pi f_D}\dfrac{1}{\sqrt{1-(|f-f_c|/f_D)^2}} & |f-f_c|\leqslant f_D\\ 0 & \text{其他}\end{cases}\end{aligned} \tag{3-29}$$

将这个功率谱密度对频率积分就是总接收功率P_r。

75

功率谱密度反映不同多普勒频率处的功率密度，也可以把它看成是对应于随机多普勒频率的概率密度函数。图3-6中的功率谱密度在$f=\pm f_D$处为无限大，因此$S_r(f)$在$f=\pm f_c\pm f_D$处也为无限大，这在实际中当然是不可能的。出现这一结果的原因在于均匀散射模型仅仅是一种近似。不过对于散射体密集的传播环境，实际的功率谱密度一般在最大多普勒频率附近达到最大值。直观来说，这个功率谱密度形状源于余弦函数的特性，而功率谱密度又可对应于随机多普勒频率$f_D(\theta)$的概率密度函数。要理解这一点，就必须认识到均匀散射假设了环境中有大量散射分量以均匀的角度和相同的平均功率到达接收机。随机选择一条路径，其到达角θ是在$[0, 2\pi]$上均匀分布的随机变量。从θ的概率密度函数可以求出随机多普勒频率$f(\theta)$的概率密度函数$p_{f_\theta}(f)$。由定义可得$p_{f_\theta}(f)$正比于多普勒频率为f处散射分量的密度，因此$S_{r_I}(f)$也正比于该密度，故可从$p_{f_\theta}(f)$得到功率谱密度。图3-7左边画出了$f_D(\theta)=f_D\cos\theta=v/\lambda\cos\theta$和它的近似折线$\underline{f}_D(\theta)$（图中虚线），右边画出

了$S_{r_I}(f)$以及对应于$\underline{f}_D(\theta)$的近似折线$\underline{S}_{r_I}(f)$。我们注意到，在一个较大的θ值范围内，近似满足$\cos(\theta)\approx\pm1$。到达角在这个范围内的多径分量的多普勒频率$f_D(\theta)\approx\pm f_D$。从功率谱上看，这些分量的功率都集中在$f\approx f_D$处。对应到图3-7中就是，左图近似折线的$\underline{f}_D(\theta)=\pm f_D$对应右图中$\pm f_D$处的冲激。而在左图中的具有固定斜率的斜线部分，对应为右图中的水平直线，每一个多径分量的功率对应一个角度增量区间。以上解释了均匀散射环境下$S_{r_I}(f)$的形状。

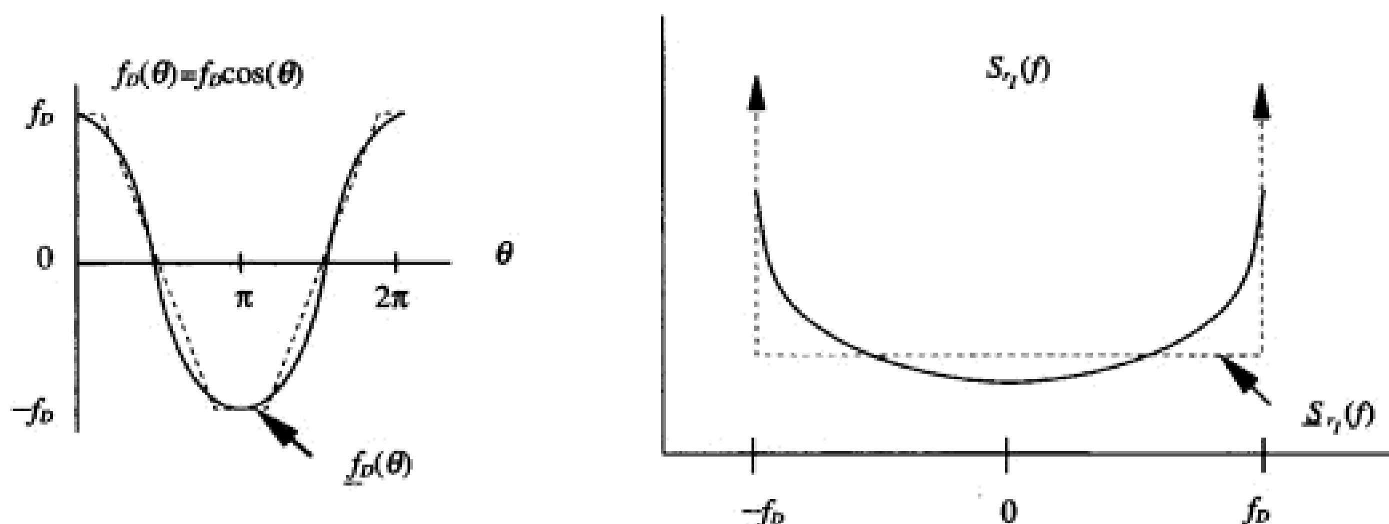


图3-7 余弦函数及功率谱密度的折线近似

典型的微蜂窝和室内环境并不是均匀散射环境，其自相关函数和功率谱密度请参见文献[5]的第1章和文献[7]的第2章。

76

功率谱密度在衰落过程的仿真中非常有用。窄带衰落过程的仿真经常是将两个独立的功率谱密度为$N_0/2$的高斯白噪声分别通过低通滤波器，该滤波器的频率响应$H(f)$须满足

$$S_{r_I}(f)=S_{r_Q}(f)=\frac{N_0}{2}|H(f)|^2 \tag{3-30}$$

其中$N_0/2$是输入白噪声的功率谱密度。滤波器的两个输出分别对应窄带衰落的同相分量和正交分量，其功率谱密度分别为$S_{r_I}(f)$和$S_{r_Q}(f)$。也可用离散滤波器以类似的方法产生

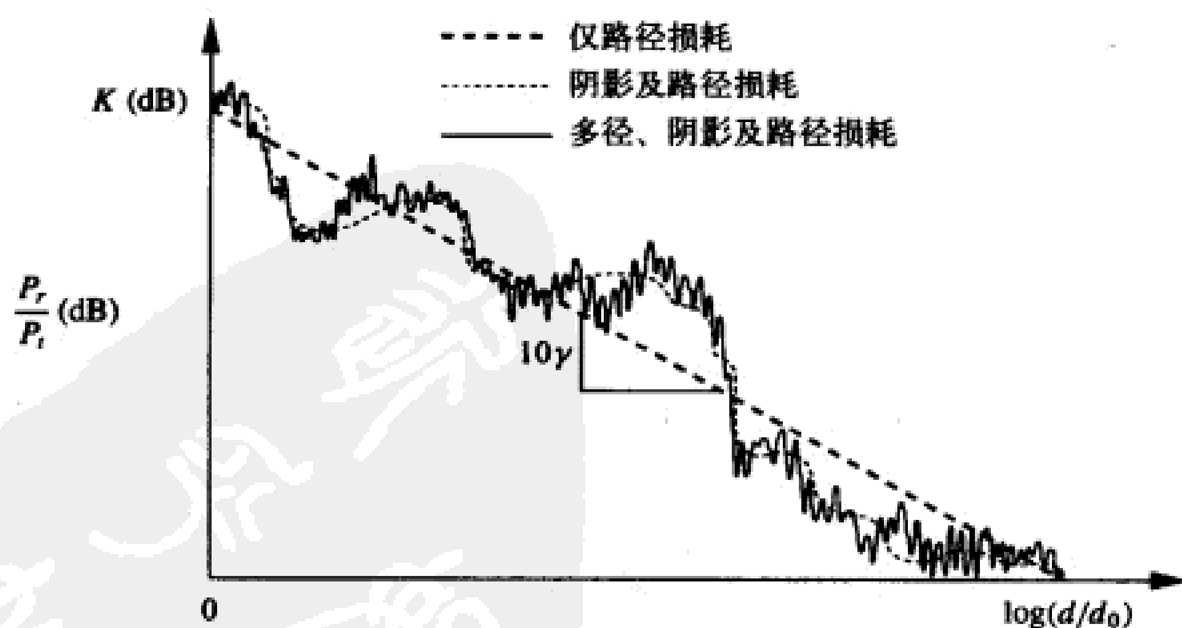


图3-8 路径损耗、阴影衰落和窄带衰落的结合

离散的衰落过程。多数通信仿真软件，如Matlab、COSSAP等，都有用这种方法实现的窄带衰落标准模块。关于仿真的具体细节及其他的方法，请参见文献[1, 7, 8]。

对于窄带无线信道，我们已经建立了三种功率随距离变化的模型，如图3-8所示。其中路径损耗和阴影衰落的模型是第2章建立的，本章增加了窄带衰落模型。由图3-8可观察到路径损耗使接收功率按d^{γ}的速度下降，γ为路径损耗指数；阴影衰落引起的功率变化比路径损耗更快，它在不相关距离X_c数量级上变化；多径衰落引起的变化最快，它在半个波长数量级上变化。将图3-8中的一小段进行水平放大，那么局部的路径损耗和阴影衰落就可看作是固定值。图3-9示出了这样放大后的一个局部，其中横坐标是线性距离$d=vt$，纵坐标是按0dBm归一化了的接收功率P_r。此图体现了一个移动台以恒速度v移动时，接收功率所呈现的平稳遍历变化。

77

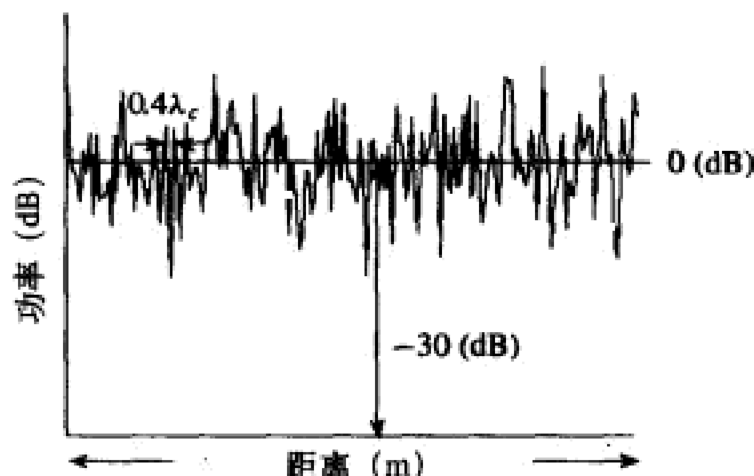


图3-9 窄带衰落

3.2.2 包络和功率分布

设X、Y是任意两个独立同分布的零均值方差为σ^2的高斯随机变量，则$Z=\sqrt{X^2+Y^2}$服从瑞利分布，Z^2服从指数分布。前面已经指出，对于均匀分布的$\phi_n(t)$，r_I和r_Q是独立同分布的零均值高斯随机变量，假定同相分量和正交分量的方差均为σ^2，则接收信号的包络为

$$z(t)=|r(t)|=\sqrt{r_I^2(t)+r_Q^2(t)} \tag{3-31}$$

它服从瑞利分布，其概率密度函数为

$$p_Z(z)=\frac{2z}{\bar{P}_r}\exp\left[-\frac{z^2}{\bar{P}_r}\right]=\frac{z}{\sigma^2}\exp\left[-\frac{z^2}{2\sigma^2}\right]\quad z\geqslant 0 \tag{3-32}$$

式中的$\bar{P}_r=\sum_n E[\alpha_n^2]=2\sigma^2$是信号的平均接收功率，它只考虑路径损耗和阴影效应，不考虑多径时的接收功率。

将$z^2(t)=|r(t)|^2$代入式（3-32）可得功率的概率密度函数为

$$p_{Z^2}(x)=\frac{1}{\bar{P}_r}\mathrm{e}^{-x/\bar{P}_r}=\frac{1}{2\sigma^2}\mathrm{e}^{-x/2\sigma^2}\quad x\geqslant 0 \tag{3-33}$$

即接收信号功率服从均值为$2\sigma^2$的指数分布。$r(t)$的复包络是$r_{LP}(t)=r_I(t)+\mathrm{j}r_Q(t)$，其相位

为$\theta=\arctan(r_Q(t)/r_I(t))$。当$r_I(t)$和$r_Q(t)$是不相关的高斯随机变量时，可证$\theta$为均匀分布，且与$|r_{LP}(t)|$独立。因此，$r(t)$的幅度为瑞利分布、相位为均匀分布，且幅度和相位彼此间相互独立。

78

例3.2：设瑞利衰落信道的平均接收功率为$\bar{P}_r=20\text{dBm}$，求接收功率低于10dBm的概率。

解：$\bar{P}_r=20\text{dBm}=100\text{mW}$，要求的是$Z^2<10\text{dBm}=10\text{mW}$的概率，即

$$p(Z^2<10)=\int_0^{10}\frac{1}{100}\mathrm{e}^{-x/100}\mathrm{d}x=0.095$$

当信道中存在一个固定的直射分量时，$r_I(t)$和$r_Q(t)$的均值不再是零。此时的接收信号是复高斯分量和直射分量的叠加，其包络服从下式给出的莱斯分布[9]

$$p_Z(z)=\frac{z}{\sigma^2}\exp\left[\frac{-(z^2+s^2)}{2\sigma^2}\right]I_0\left(\frac{zs}{\sigma^2}\right)\quad z\geqslant 0 \tag{3-34}$$

上式中$s^2=\alpha_0^2$是直射分量的功率，$2\sigma^2=\sum_{n,n\neq 0}E[\alpha_n^2]$是其他非直射分量的平均功率。$I_0$是修正的零阶贝塞尔函数。莱斯衰落的平均接收功率为

$$\bar{P}_r=\int_0^{\infty}z^2p_Z(z)\,\mathrm{d}z=s^2+2\sigma^2 \tag{3-35}$$

常用衰落参数K来表示莱斯分布，其定义为

$$K=\frac{s^2}{2\sigma^2} \tag{3-36}$$

K代表直射和非直射分量的功率比。$K=0$就是瑞利衰落，$K=\infty$就是无衰落（信道没有多径成分，只有固定的直射分量）。衰落参数K反映了衰落的严重程度，K越小表示衰落越严重，K越大表示衰落越轻。经过变量代换$s^2=K\bar{P}_r/(K+1)$和$2\sigma^2=\bar{P}_r/(K+1)$后，可以用$K$、$\bar{P}_r$为参数给出莱斯分布为

$$p_Z(z)=\frac{2z(K+1)}{\bar{P}_r}\exp\left[-K-\frac{(K+1)z^2}{\bar{P}_r}\right]I_0\left(2z\sqrt{\frac{K(K+1)}{\bar{P}_r}}\right)\quad z\geqslant 0 \tag{3-37}$$

瑞利分布和莱斯分布都能用数学方法从所假设的物理信道模型中导出[3, 9]。不过，有些实验数据和这两个分布都不太吻合。因此，人们提出了一个能吻合许多不同实验数据的更为通用的衰落分布，它就是下式给出的Nakagami衰落分布

$$p_Z(z)=\frac{2m^m z^{2m-1}}{\Gamma(m)\bar{P}_r^m}\exp\left[\frac{-mz^2}{\bar{P}_r}\right]\quad m\geqslant 0.5 \tag{3-38}$$

式中$\bar{P}_r$为平均接收功率，$\Gamma(.)$为伽马函数。Nakagami分布有两个参数：$\bar{P}_r$和m，$\bar{P}$为平均接收功率，m为衰落参数。$m=1$时，式（3-38）退化为瑞利衰落。令$m=(K+1)^2/(2K+1)$，则式（3-38）近似于参数为K的莱斯分布。$m=\infty$代表无衰落，此时$Z=\sqrt{\bar{P}_r}$是一个常数。因此Nakagami分布不仅可以表示瑞利和莱斯分布，还能表示其他许多衰落。有些实验数据对应的m参数小于1，这样的Nakagami衰落对系统性能造成的恶化比瑞利衰落还要严重。

通过变量代换可得到Nakagami衰落的功率分布为

$$p_{Z^2}(x) = \left(\frac{m}{\bar{P}_r}\right)^m \frac{x^{m-1}}{\Gamma(m)} \exp\left[\frac{-mx}{\bar{P}_r}\right] \tag{3-39}$$

3.2.3 电平通过率和平均衰落时长

电平通过率L_Z定义为单位时间内信号包络向下穿过电平Z的平均次数。计算L_Z需要信号包络$z=|r|$和z的时间导数$\dot{z}$的联合分布$p(z, \dot{z})$。下面我们基于这个联合分布推导L_Z的表达式。

79

考虑图3-10所示的衰落过程。在时间间隔dt内，信号包络在区间（Z, Z+dz）内、包络斜率在区间$[\dot{z}, \dot{z}+d\dot{z}]$内的平均时间为$A=p(Z, \dot{z})dzd\dot{z}dt$。对于给定的斜率$\dot{z}$，包络穿过区间（Z, Z+dz）一次所需的时间为$B=dz/\dot{z}$。比值$A/B=\dot{z}p(Z, \dot{z})d\dot{z}dt$是包络z在时间间隔d$t$内，以斜率$\dot{z}$穿过区间（Z, Z+dz）的平均次数。因此，在时间间隔[0, T]内，包络以区间$[\dot{z}, \dot{z}+d\dot{z}]$内的斜率向下通过Z的平均次数为

$$\int_0^T \dot{z}p(Z,\dot{z})\,d\dot{z}\,dt = \dot{z}p(Z,\dot{z})\,d\dot{z}T \tag{3-40}$$

进一步得到，在时间间隔[0, T]内，包络以负斜率穿过电平Z的平均次数为

$$N_Z = T\int_{-\infty}^0 \dot{z}p(Z,\dot{z})\,d\dot{z} \tag{3-41}$$

最终，一秒内包络穿过电平Z的平均次数，也即电平通过率为

$$L_Z = \frac{N_Z}{T} = \int_{-\infty}^0 \dot{z}p(Z,\dot{z})\,d\dot{z} \tag{3-42}$$

这是一个通用结果，适用于任何随机过程。

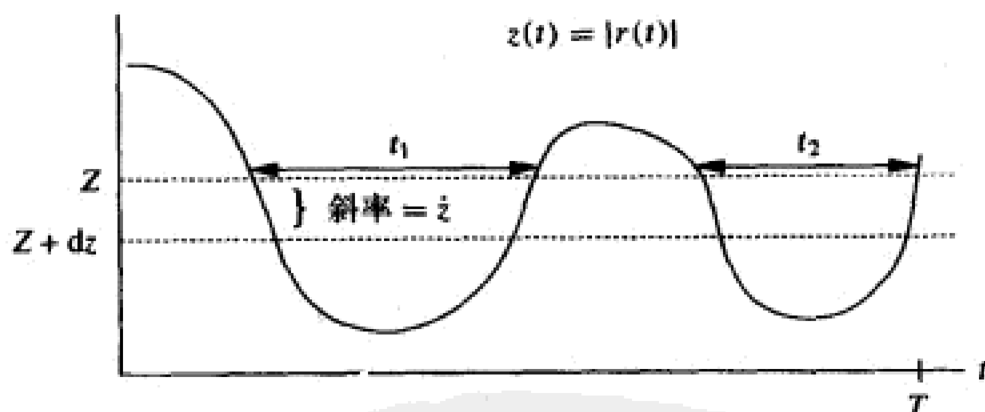


图3-10 衰落过程的电平通过率和衰落时长

莱斯衰落下z和$\dot{z}$的联合概率密度函数见文献[9]及[7]的2.1节。将其代入式（3-42）便得到莱斯衰落的电平通过率为

$$L_Z = \sqrt{2\pi(K+1)}f_D\rho e^{-K-(K+1)\rho^2} I_0\left(2\rho\sqrt{K(K+1)}\right) \tag{3-43}$$

式中$\rho=Z/\sqrt{\bar{P}_r}$。容易证明接收信号功率穿过门限值γ_0的通过率也是式（3-43），只是需重新定义$\rho=\sqrt{\gamma_0/\sqrt{\bar{P}_r}}$。对于瑞利衰落（$K=0$），电平通过率简化为

$$L_Z = \sqrt{2\pi}f_D\rho e^{-\rho^2} \tag{3-44}$$

[80] 式中$\rho = Z/\sqrt{\overline{P_r}}$。

平均信号衰落时长定义为信号包络值低于给定的目标电平Z的平均时间。一般根据特定的性能指标（如误比特率）所需要的接收功率或幅度来确定这个目标电平值，如果信号幅度或者功率低于目标值，就称系统处于中断状态。令t_i表示在时间间隔[0, T]内第i次低于电平Z的持续时长（见图3-10），也就是第i次向下通过Z后，继续停留在Z以下的时长。$z(t)$是平稳遍历过程，故对于足够大的T我们有

$$p(z(t) < Z) = \frac{1}{T}\sum_i t_i \tag{3-45}$$

于是，无限大时间内的平均衰落时长为

$$\bar{t}_Z = \frac{1}{TL_Z}\sum_{i=1}^{L_Z T} t_i \approx \frac{p(z(t) < Z)}{L_Z} \tag{3-46}$$

对$p(z(t)<Z)$使用瑞利分布得到

$$\bar{t}_Z = \frac{e^{\rho^2} - 1}{\rho f_D \sqrt{2\pi}} \tag{3-47}$$

其中$\rho = Z/\sqrt{\overline{P_r}}$。式（3-47）是对目标幅度值为$Z$、平均幅度值为$\sqrt{\overline{P_r}}$的信号幅值而言的平均衰落时长。通过变量代换容易得知，取$\rho = \sqrt{P_0/\overline{P_r}}$后，式（3-47）也是相对于目标功率为$P_0$、平均功率为$\overline{P_r}$的信号功率而言的平均衰落时长。式（3-47）中平均衰落时长随多普勒频移的增大而减小，这是因为信道变化越快时，它每次停留在给定衰落电平之下的时间也越短。当$\rho \gg 1$时平均衰落时长还随ρ的增大而增加，这是因为目标电平值越高时，接收信号低于目标值的可能性越大。莱斯衰落的平均衰落时长计算比较复杂，参见文献[7]的1.4节。

平均衰落时长可以反映出受深衰落影响的比特或者符号个数。考虑一个比特间隔为T_b的未编码系统，假设$z<Z$时误比特率很高。如果$T_b \approx \bar{t}_Z$，那么系统中的出错的那个比特（对应$z<Z$）之前和之后的比特很可能是正确的，因为对于前后的比特必然有$z>Z$，故发生单比特错误。如果$T_b \ll \bar{t}_Z$，那么错误比特之后接收到的比特也满足$z<Z$，因此可能发生长串突发错误。而当$T_b \gg \bar{t}_Z$时，解调器中的积分运算将把衰落平均掉，此时衰落的影 [81] 响可以忽略。第8章介绍编码和交织时将进一步讨论这一问题。

例3.3：某话音系统中，当接收信号功率大于等于平均接收功率的一半时，其误码率值是可接受的。如果误比特率低于可接受值的持续时间超过120ms，通话就会被中止。今若要求话音质量不可接受的平均持续时间小于$t=60$ms，求瑞利衰落信道下相应的多普勒频移范围。

解：目标接收功率是平均值的一半，故$P_0 = 0.5\overline{P_r}$，$\rho = \sqrt{0.5}$。要求

$$\bar{t}_Z = \frac{e^{0.5} - 1}{f_D\sqrt{\pi}} \leqslant t = 0.06$$

即要求$f_D \geqslant (e-1)/\left(0.06\sqrt{2\pi}\right) = 6.1\ \text{Hz}$。

3.2.4 有限状态马氏信道

对于无线系统的性能分析来说，前面给出的平衰落模型过于复杂，我们需要一些更简单但又能反映平衰落信道主要特性的模型。有限状态马氏信道（finite-state Markov channel，FSMC）就是这样的简化模型之一。FSMC将时间取样后的信道近似为离散时间的马尔科夫过程，采样间隔一般等于码元周期。所有可能的衰落增益构成一个有限的信道状态集合。每隔T时间，信道按照马尔科夫转移概率在这些状态之间转移。有限状态马氏模型可用于近似卫星信道[10]、室内信道[11]、瑞利衰落信道[12-13]、莱斯衰落信道[14]和Nakagami-m衰落信道[15]等的数学模型或实验模型，也可用于系统设计和系统性能分析[13,16]。对于系统性能分析来说，一阶FSMC模型往往不够精确，因而广泛使用的是高阶FSMC模型。FSMC模型一般用来描述衰落信道的幅度变化，也有人把它用于描述衰落的相位[17]或信道的相位噪声[18]的变化。

文献[12]中建立了瑞利衰落的FSMC模型。它把瑞利衰落中的时变信噪比γ（$0\leqslant\gamma\leqslant\infty$）量化为许多区间，第$j$个区间$R_j$对应一个信噪比范围$R_j=\{\gamma: A_j\leqslant\gamma\leqslant A_{j+1}\}$，区间的边界$A_j$和总的区间个数是模型中的参数。模型假定在时间间隔$T$内$\gamma$值保持在同一个区间内，在这个时间间隔结束时，可能继续停留在本区间内或者转移到相邻的信噪比区间，即若某个时刻的信道状态是R_j，则下一时刻的信道状态只可能转移到R_{j-1}、R_j或者R_{j+1}，不能到达别的状态，这在f_DT很小时是一个合理的假设。在这样的假设下，文献[12]给出了区间之间的转移概率为

$$p_{j,j+1}=\frac{L_{j+1}T}{\pi_j}\quad p_{j,j-1}=\frac{L_jT}{\pi_j}\quad p_{j,j}=1-p_{j,j+1}-p_{j,j-1} \tag{3-48}$$

其中L_j是A_j处的电平通过率，π_j是第j个区间的稳态分布：$\pi_j=p(\gamma\in R_j)=p(A_j\leqslant\gamma\leqslant A_{j+1})$。 82

3.3 宽带衰落模型

宽带信号中多增加了一个失真因素，就是多径时延扩展。若发送脉冲的持续时间为T，到接收端脉冲就会扩展成（$T+T_m$），T_m为多径时延扩展，如图3-11所示，脉宽为T的脉冲信号经过多径信道发送。第5章将要讲到，线性调制由一串脉冲组成，每个脉冲通过幅度和相位携带数据比特或符号[1]。如果多径时延扩展$T_m<<T$，那么如图右上所示，接收到的多径分量在时间上大致重叠在一起。这些重叠的信号相互干涉能造成窄带衰落，但脉冲在时间上没有明显的扩展，因此对后续脉冲的干扰很小。而当多径时延扩展$T_m>>T$时，那么如图右下所示，每一径都是可分辨的。这些多径分量将对后续脉冲（图中虚线表示的脉冲）造成干扰，这种现象称为*码间干扰*（inter-symbol interference，ISI）。

对于多径时延扩展造成的这种失真，存在一些对抗的方法，包括均衡、多载波调制和扩频等，它们将在第11章~第13章中介绍。如果我们把系统设计成$T>>T_m$，则不需要采取措施来对抗码间干扰，但这会限制数据的传输速率。多载波调制和扩频技术通过改变发送信号的特性来最大限度地避免码间干扰，但它们依然会受到频率选择性衰落影响，

1. 出于频段利用率的原因，线性调制一般不采用矩形脉冲，详见5.4节。

详见3.3.2节。

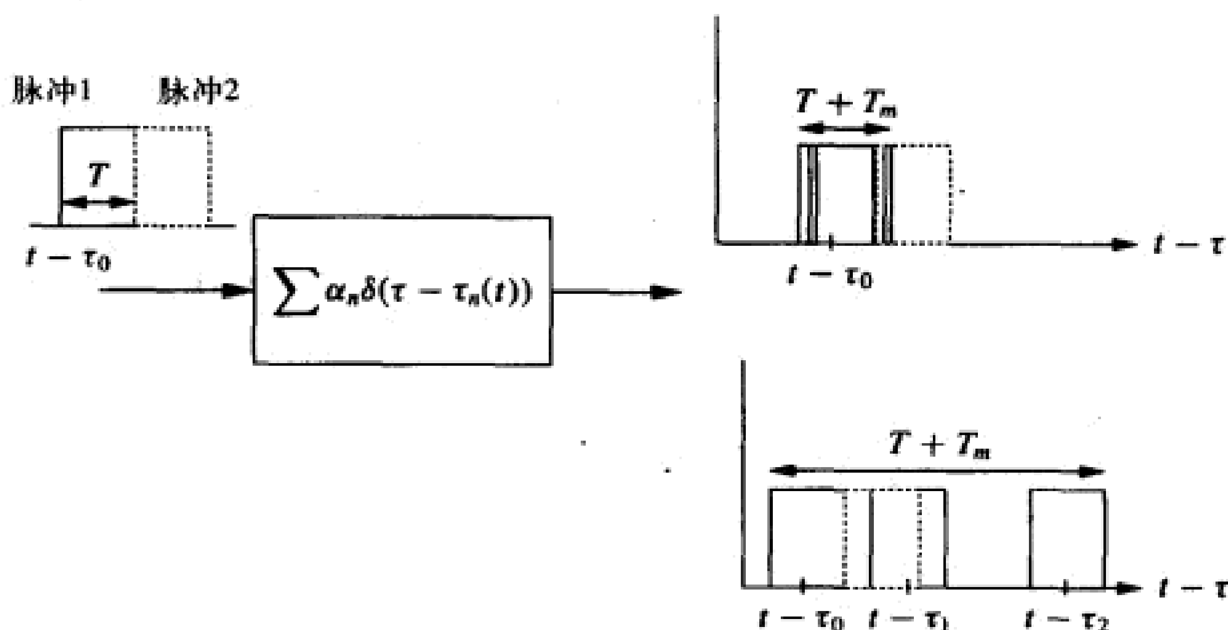


图3-11 多径的分辨性

宽带模型有别于窄带衰落模型的是，随着发送信号带宽B逐步增加到$T_m \approx B^{-1}$的时候，近似关系$u(t-\tau_n(t)) \approx u(t)$不再成立。这样，接收信号是原发送信号经过不同时延$\tau_n$、不同相移$\phi_n(t)$之后各分量的和。相位差会造成的信号相互抵消，而$u(t-\tau_n)$和$u(t)$的时延差会造成失真。

[83]

宽带衰落虽然不再近似满足式（3-11），但径数很大且每一径相位都服从均匀分布时，接收信号仍然是一个幅度为瑞利分布的零均值复高斯过程，与窄带衰落不同之处是多径分量之间的时延差。窄带信号的多径时延差小于信号带宽的倒数，这使得式（3-6）中多径合并的结果只是给原发信号增加了一个体现随机幅度和相位的随机过程。我们在3.2节中用自相关函数（或功率谱密度）和一维分布来描述这个随机过程。对于宽带信号，仅仅反映幅度和相位随机过程是不够的，因为不同多径分量的时延扩展将会使接收信号产生失真，所以必须还要同时考虑时延扩展和信道时变性对信号的影响。

我们从信道的等效基带时变冲激响应$c(\tau, t)$来着手考虑宽带信道。先假定$c(\tau, t)$对τ和t都是确定的连续函数[1]。注意τ代表给定多径时延下的冲激响应，t代表时变性。对$c(\tau, t)$进行关于t的傅里叶变换得

$$S_c(\tau,\rho) = \int_{-\infty}^{\infty} c(\tau,t)\mathrm{e}^{-\mathrm{j}2\pi\rho t}\,\mathrm{d}t \tag{3-49}$$

称$S_c(t, \rho)$为等效低通冲激响应$c(\tau, t)$的**确定散射函数**（deterministic scattering function）。由于它是$c(\tau, t)$关于时变参数t的傅里叶变换，因此散射函数$S_c(\tau, \rho)$能通过频率参数ρ反映了信道的多普勒特性。

一般来说，多径分量的幅度、相位、时延以及径数都是随机的，所以式（3-6）给出的时变信道冲激响应$c(\tau, t)$不是确定的而是随机的。这就需要用统计的方法或者用测量的方法来进行研究。当径数很大时可以借助中心极限定理把$c(\tau, t)$假定为复高斯过程。

1. 本节对宽带信道的研究方法也可推广到离散时间的信道（τ离散）中。需将积分运算改为求和运算，将傅里叶变换改为离散傅里叶变换。

这样，信道的统计特性完全由其同相分量和正交分量的均值、自相关函数、互相关函数确定。如同窄带信道的情形一样，在宽带信道中我们假定每径的相位服从均匀分布，则$c(\tau, t)$的同相分量和正交分量也是相互独立的高斯过程，它们的自相关函数相同、均值为零、互相关为零。这个结果在径数较少时也能成立，只要各径的幅度为瑞利分布、相位为均匀分布。但请注意，如果信道中存在一个起主导作用的直射径，这些结论并不成立。

84

因此，$c(\tau, t)$的统计特性就由其自相关函数来确定，定义为

$$A_c(\tau_1, \tau_2; t, t+\Delta t) = E[c^*(\tau_1; t)c(\tau_2; t+\Delta t)] \tag{3-50}$$

实际信道大多是广义平稳的，所以在时刻t和（$t+\Delta t$）所测量的联合统计特性只与时间差Δt有关。广义平稳信道的带通响应$h(\tau, t)=\mathrm{Re}\{c(\tau, t)e^{j2\pi f_c t}\}$的自相关函数可以从$A_c(\tau_1, \tau_2; t, t+\Delta t)$导出为[19]$A_h(\tau_1, \tau_2; t, t+\Delta t)=0.5\mathrm{Re}\{A_c(\tau_1, \tau_2; t, t+\Delta t)e^{j2\pi f_c \Delta t}\}$[1]。我们假定信道模型为广义平稳，其自相关函数与t无关：

$$A_c(\tau_1, \tau_2; \Delta t) = E[c^*(\tau_1; t)c(\tau_2; t+\Delta t)] \tag{3-51}$$

此外，实际环境中时延分别为τ_1和$\tau_2 \neq \tau_1$的两个径来自不同的散射体，它们的信道响应是不相关的。称这样的信道为*不相关散射*（uncorrelated scattering，US）信道。如果一个信道同时具备广义平稳（WSS）和不相关散射（US）这两个特性，则称为*广义平稳不相关散射*（WSSUS）信道。WSSUS模型最早是由Bello在文献[19]中提出的，他在这篇文献中还建立了自相关函数的二维变换关系。这些变换关系将在3.3.4节中讨论。把不相关散射性质加入式（3-51）得到

$$E[c^*(\tau_1; t)c(\tau_2; t+\Delta t)] = A_c(\tau_1; \Delta t)\delta[\tau_1-\tau_2] \triangleq A_c(\tau; \Delta t) \tag{3-52}$$

式中$A_c(\tau; \Delta t)$的是信道的平均输出功率，它是多径时延$\tau=\tau_1=\tau_2$和观察间隔Δt的函数。该函数假设$\tau_1 \neq \tau_2$时，τ_1和τ_2满足$|\tau_1-\tau_2|>B^{-1}$，若不然，这两个径就是不可分辨的，应视为一个合一的径，其时延为$\tau \approx \tau_1 \approx \tau_2$。

随机信道的*散射函数*（scattering function）定义为$A_c(\tau; \Delta t)$关于Δt的傅里叶变换：

$$S_c(\tau, \rho) = \int_{-\infty}^{\infty} A_c(\tau, \Delta t)e^{-j2\pi\rho\Delta t}\, d\Delta t \tag{3-53}$$

它反映信道的平均输出功率作为多径时延τ和多普勒频率ρ的函数。注意我们对确定散射函数和随机散射函数都用记号$S_c(\tau, \rho)$来表示，可以根据信道冲激响应是确定或随机来确定$S_c(\tau, \rho)$是确定散射函数还是随机散射函数，所以不会造成混淆。图3-12给出了一个典型的散射函数。

下面几节将讨论宽带信道最重要的一些特征，包括功率时延谱、相干带宽、多谱勒

1. 容易证明，带通信道响应$h(\tau, t)$的自相关函数为$E[h(\tau_1, t)h(\tau_2, t+\Delta t)]=\frac{1}{2}\mathrm{Re}\{A_c(\tau_1, \tau_2; t, t+\Delta t)e^{j2\pi f_c \Delta t}\}+\frac{1}{2}\mathrm{Re}\{\hat{A}_c(\tau_1, \tau_2; t, t+\Delta t)e^{j2\pi f_c(2t+\Delta t)}\}$。其中$\hat{A}_c(\tau_1, \tau_2; t, t+\Delta t)=E[c(\tau_1, t)c(\tau_2, t+\Delta t)]$。由$c(\tau, t)$是广义平稳过程可得$\hat{A}_c(\tau_1, \tau_2; t, t+\Delta t)=0$。于是$E[h(\tau_1, t)h(\tau_2, t+\Delta t)]=\frac{1}{2}\mathrm{Re}\{A_c(\tau_1, \tau_2; t, t+\Delta t)e^{j2\pi f_c \Delta t}\}$。

85 功率谱和相关时间，这些特征都是从自相关函数$A_c(\tau, t)$或散射函数$S_c(\tau, \rho)$导出的。

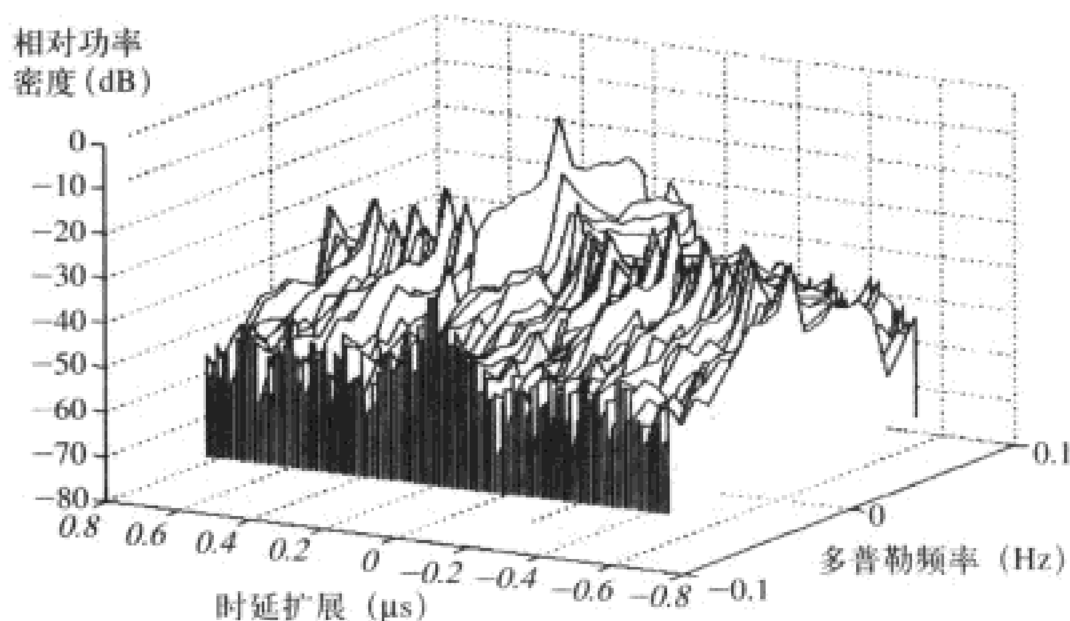


图3-12 散射函数

3.3.1 功率时延谱

令式（3-52）中的$\Delta t=0$就定义了*功率时延谱*（power delay profile）$A_c(\tau)$，即$A_c(\tau)\triangleq A_c(\tau,0)$，也称为*多径强度谱*（multipath intensity profile）。功率时延谱表示给定多径时延处的平均功率，很容易测量。一般用功率时延谱来定义*平均时延扩展*μ_{T_m}和*均方根时延扩展*σ_{T_m}：

$$\mu_{T_m}=\frac{\int_0^\infty \tau A_c(\tau)\,\mathrm{d}\tau}{\int_0^\infty A_c(\tau)\,\mathrm{d}\tau} \tag{3-54}$$

$$\sigma_{T_m}=\sqrt{\frac{\int_0^\infty(\tau-\mu_{T_m})^2 A_c(\tau)\,\mathrm{d}\tau}{\int_0^\infty A_c(\tau)\,\mathrm{d}\tau}} \tag{3-55}$$

注意对于所有τ均有$A_c(\tau)\geqslant 0$，故若定义随机时延扩展T_m的概率密度函数为

$$p_{T_m}(\tau)=\frac{A_c(\tau)}{\int_0^\infty A_c(\tau)\,\mathrm{d}\tau} \tag{3-56}$$

则μ_{T_m}和σ_{T_m}分别是T_m在此分布下的均值和均方根值。无论是用式（3-56）的分布还是直接用式（3-54）和（3-55）来定义，它们都是把各径的时延按其功率进行了加权，因此弱径对时延扩展的贡献比强径小。低于背景噪声的径不会对时延扩展产生明显的影响。86

如果τ超过某个时延T后，$A_c(\tau)$值近似为零，则可用这个T值来粗略反映信道的时延扩展，一般取为均方根时延扩展的整倍数。例如，假设$\tau>3\sigma_{T_m}$时$A_c(\tau)\approx 0$，若线性调制的码元周期是T_s，则当$T_s<<\sigma_{T_m}$时存在严重的码间干扰，而当$T_s>>\sigma_{T_m}$时码间干扰可以忽略。计算时可将$T_s<\sigma_{T_m}/10$视同$T_s<<\sigma_{T_m}$，将$T_s>10\sigma_{T_m}$视同$T_s>>\sigma_{T_m}$。当T_s的大小和σ_{T_m}在同一个数量级时，系统中将会有一些码间干扰，对系统性能是否有显著影响取决于具体的系统

和信道。在后面的章节中我们将研究码间干扰对线性调制系统的影响以及对抗方法。

许多有大量反射体的信道近似有$\mu_{T_m} \approx \sigma_{T_m}$，但并非所有信道都如此，$\mu_{T_m}$和$\sigma_{T_m}$的准确关系取决于$A_c(\tau)$的形状。若信道中没有直射径、多径数量少、且各径的时延差不多一样大时，$\mu_{T_m} >> \sigma_{T_m}$。这种情况下，用这个比较大的$\mu_{T_m}$值作为时延扩展的度量容易产生误导。实际上，发送信号的所有分量大致同时到达，接收端解调器将会同步到这个公共的时延上。同步电路一般会锁定在平均时延上，此时均方根时延扩展真正反映了信道的时间扩展特性。

例3.4： 常见的功率时延谱模型是单边指数分布：

$$A_c(\tau) = \frac{1}{\bar{T}_m} e^{-\tau/\bar{T}_m} \quad \tau \geqslant 0$$

请证明平均时延扩展 $\mu_{T_m} = \bar{T}_m$，并求均方根时延扩展。

解： 易证$A_c(\tau)$的积分为1，因此平均时延扩展为

$$\mu_{T_m} = \frac{1}{\bar{T}_m} \int_0^{\infty} \tau e^{-\tau/\bar{T}_m} \mathrm{d}\tau = \bar{T}_m$$

均方根时延扩展为

$$\sigma_{T_m} = \sqrt{\frac{1}{\bar{T}_m} \int_0^{\infty} \tau^2 e^{-\tau/\bar{T}_m} \mathrm{d}\tau - \mu_{T_m}^2} = 2\bar{T}_m - \bar{T}_m = \bar{T}_m$$

可见，对于指数分布的功率时延谱，平均时延扩展等于均方根时延扩展。

87

例3.5： 某宽带信道的多径强度谱为

$$A_c(\tau) = \begin{cases} e^{-\tau/0.00001} & 0 \leqslant \tau \leqslant 20\mu s \\ 0 & \text{其他} \end{cases}$$

请计算信道的平均时延扩展和均方根时延扩展。若有线性调制信号通过此信道传送，求不发生码间干扰的最大符号速率。

解： 平均时延扩展为

$$\mu_{T_m} = \frac{\int_0^{20\cdot 10^{-6}} \tau e^{-\tau/0.00001} \mathrm{d}\tau}{\int_0^{20\cdot 10^{-6}} e^{-\tau/0.00001} \mathrm{d}\tau} = 6.87\mu s$$

均方根时延扩展为

$$\sigma_{T_m} = \sqrt{\frac{\int_0^{20\cdot 10^{-6}} (\tau - \mu_{T_m})^2 e^{-\tau/0.00001} \mathrm{d}\tau}{\int_0^{20\cdot 10^{-6}} e^{-\tau/0.00001} \mathrm{d}\tau}} = 5.25\ \mu s$$

本例中平均时延扩展和均方根时延扩展近似相等。为了避免码间干扰，要求码元周期远大于σ_{T_m}，取$T_s > 10\sigma_{T_m}$。由此得$T_s = 52.5\mu s$，相应的符号速率为$R_s = 1/T_s = 19.04$k符号/秒。这个速率对许多无线系统来说是一个很大的限制。例如对于二进制调制来说，符号速率就是比特速率（bit/s），而话音业务要求的速率大约在32kbit/s数量级，高速数据业务要求速率为10Mbit/s～100Mbit/s。

PDG

3.3.2 相干带宽

也可以通过对$c(\tau, t)$做关于τ的傅里叶变换，在频域研究时变多径信道。傅里叶变换的结果是一个随机过程：

$$C(f;t)=\int_{-\infty}^{\infty}c(\tau;t)\mathrm{e}^{-\mathrm{j}2\pi f\tau}\,\mathrm{d}\tau \tag{3-57}$$

因为$c(\tau, t)$是以t为时间参量的零均值复高斯随机过程，所以式（3-57）的傅里叶变换是对零均值复高斯过程的求和[1]，因此$C(f; t)$也是一个零均值的复高斯过程，其特性完全由自相关函数确定。又由于$c(\tau, t)$是广义平稳过程，故$C(f; t)$也是广义平稳过程。这样，式(3-57)的自相关函数可以表示为

$$A_C(f_1,f_2;\Delta t)=E[C^*(f_1;t)C(f_2;t+\Delta t)] \tag{3-58}$$

可将其简化为如下的形式：

$$\begin{aligned}A_C(f_1,f_2;\Delta t)&=E\left[\int_{-\infty}^{\infty}c^*(\tau_1;t)\mathrm{e}^{\mathrm{j}2\pi f_1\tau_1}\,\mathrm{d}\tau_1\int_{-\infty}^{\infty}c(\tau_2;t+\Delta t)\mathrm{e}^{-\mathrm{j}2\pi f_2\tau_2}\,\mathrm{d}\tau_2\right]\\&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}E[c^*(\tau_1;t)c(\tau_2;t+\Delta t)]\mathrm{e}^{\mathrm{j}2\pi f_1\tau_1}\mathrm{e}^{-\mathrm{j}2\pi f_2\tau_2}\,\mathrm{d}\tau_1\,\mathrm{d}\tau_2\\&=\int_{-\infty}^{\infty}A_c(\tau,\Delta t)\mathrm{e}^{-\mathrm{j}2\pi(f_2-f_1)\tau}\,\mathrm{d}\tau\\&=A_C(\Delta f;\Delta t)\end{aligned} \tag{3-59}$$

式中$\Delta f=f_2-f_1$，第三个等式来自$c(\tau, t)$的广义平稳性和非相关散射性。因此，$C(f; t)$关于频率的自相关函数只与频率差Δf有关。测量$A_C(\Delta f; \Delta t)$的方法是：发送一对载频差为Δf的正弦波，然后在接收端求出这两个正弦波经过时间间隔Δt后的互相关值。

88

如果定义$A_C(\Delta f)\triangleq A_C(\Delta f; 0)$，那么由式（3-59）得

$$A_C(\Delta f)=\int_{-\infty}^{\infty}A_c(\tau)\mathrm{e}^{-\mathrm{j}2\pi\Delta f\tau}\,\mathrm{d}\tau \tag{3-60}$$

即$A_C(\Delta f)$是功率时延谱的傅里叶变换。由于$A_C(\Delta f)=E[C^*(f; t)C(f+\Delta f; t)]$是自相关函数，因此若$A_C(\Delta f)\approx 0$，则频率相隔$\Delta f$处的信道响应近似独立。若存在某个$B_c$，使得对于所有的$\Delta f>B_c$均有$A_C(\Delta f)\approx 0$，$B_c$就叫做信道的*相干带宽*（coherence bandwidth）。由$A_c(\tau)$与$A_C(\Delta f)$之间的傅里叶变换关系可知，若对$\tau>T$有$A_c(\tau)\approx 0$，则对$\Delta f>1/T$有$A_C(\Delta f)\approx 0$。因此，能使信道响应近似独立的最小频率间隔为$B_c\approx 1/T$，T一般取为$A_c(\tau)$的均方根时延扩展σ_{T_m}。更一般的近似是$B_c\approx k/\sigma_{T_m}$，$k$取决于$A_c(\tau)$的形状以及我们具体怎么定义相干带宽。例如李建业[20]指出，信道相关度大于0.9的频率范围为$B_c\approx 0.02/\sigma_{T_m}$，信道相关度大于0.5的频率范围为$B_c\approx 0.2/\sigma_{T_m}$。

总的来说，如果发送一个带宽$B<<B_c$的窄带信号，那么信号带宽内的衰落是高度相关的，也就是说整个带宽上的衰落近似相等，通常称此为*平衰落*（flat fading）。另一方

1. 积分是无限求和。

PDG

面，如果信号带宽$B>>B_c$，那么在间隔超过相干带宽的两个频率点上的信道幅度值近似独立。此时信号带宽内信道幅度变化很大，这样的衰落称为频率选择性（frequency selective）衰落。当$B\approx B_c$时的情形介于平衰落和频率选择性衰落之间。线性调制的信号带宽B与码元间隔T_s成反比，平衰落对应于$T_s\approx 1/B>>1/B_c\approx\sigma_{T_m}$，此时码间干扰可以忽略。频率选择性衰落对应于$T_s\approx 1/B<<1/B_c=\sigma_{T_m}$，此时会有严重的码间干扰。虽然多载波调制和扩频技术能减少码间干扰，但它们是宽带的信号方式，因而整个信号带宽内也会发生频率选择性衰落，这也将降低系统的性能，第12章和第13章将对此进行详细的讨论。

89

图3-13示出了功率时延谱$A_c(\tau)$和它的傅里叶变换$A_C(\Delta f)$。在$A_C(\Delta f)$的图中还同时画出了一个带宽远小于B_c的窄带信号和一个带宽远大于B_c的宽带信号。在窄带信号的带宽内，自相关函数$A_C(\Delta f)$是平的，信号将经历平衰落，码间干扰可以忽略。在宽带信号的带宽内，自相关函数$A_C(\Delta f)$逐渐降为零，说明信号带宽内不同部分的衰落是相互独立的，因此是频率选择性衰落，在这个信道上传送的线性调制信号会经历严重的码间干扰。

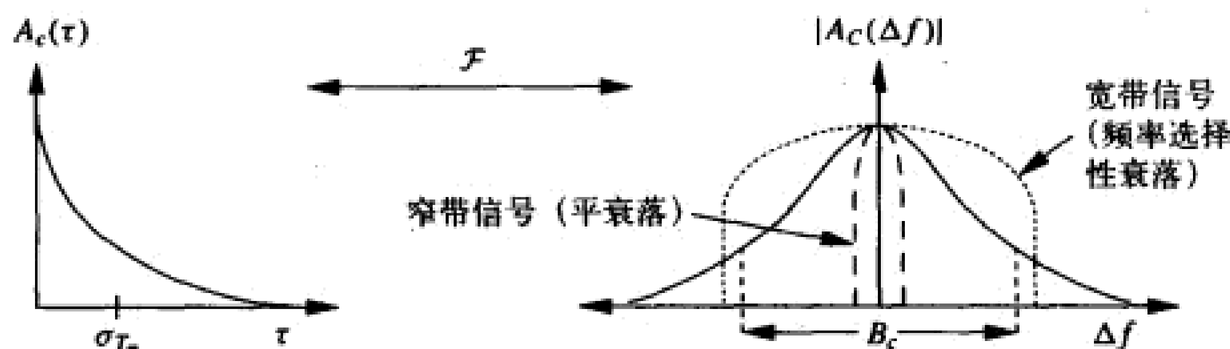


图3-13 功率时延谱、均方根时延扩展及相干带宽

例3.6：室内信道中$\sigma_{T_m}\approx 50\text{ns}$，室外微小区$\sigma_{T_m}\approx 30\mu\text{s}$。假设发送线性调制信号，求这两种环境中，能使ISI可以忽略的最大符号速率$R_s=1/T_s$。

解：码间干扰可以忽略要求的条件是$T_s>>\sigma_{T_m}$，也即$T_s\geqslant 10\sigma_{T_m}$，相应的码元速率为$R_s=1/T_s\leqslant 0.1/\sigma_{T_m}$。由$\sigma_{T_m}\approx 50\text{ns}$可得到$R_s\leqslant 2\text{Mbit/s}$，由$\sigma_{T_m}\approx 30\mu\text{s}$可得到$R_s\leqslant 3.33\text{kbit/s}$。目前的室内系统速率可达到50Mbit/s，室外系统可达到2.4Mbit/s。为了使线性调制保持这样高的数据速率而不会因码间干扰造成性能严重下降，必须采取一些对抗码间干扰的措施。另外，室内系统时延扩展较小，所以码间干扰轻一些，这是室内系统比室外系统速率高的原因。

3.3.3 多普勒功率谱和信道相干时间

发射机或接收机的移动产生的信道时变性引起了多普勒频移，其特性可用$A_C(\Delta f;\Delta t)$关于Δt的傅里叶变换来反映：

$$S_C(\Delta f;\rho)=\int_{-\infty}^{\infty}A_C(\Delta f;\Delta t)\mathrm{e}^{-\mathrm{j}2\pi\rho\Delta t}\,\mathrm{d}\Delta t \tag{3-61}$$

为了表示单一频率上的多普勒特性，令Δf为零并定义$S_C(\rho)\triangleq S_C(0;\rho)$，即

$$S_C(\rho)=\int_{-\infty}^{\infty}A_C(\Delta t)\mathrm{e}^{-\mathrm{j}2\pi\rho\Delta t}\,\mathrm{d}\Delta t \tag{3-62}$$

式中$A_c(\Delta t) \triangleq A_c(\Delta f=0;\ \Delta t)$。注意$A_c(\Delta t)$是信道冲激响应沿时间的自相关函数。$A_c(\Delta t=T)=0$表示相隔时间$T$后，信道冲激响应的观察值不相关，因为信道是高斯过程，故此也是独立的。定义信道**相干时间**（channel coherence time）T_c为这样一个时间间隔Δt，此间隔内的$A_c(\Delta t)$近似不为零。这样，时变信道在大约经过T_c秒后变得不相关。函数$S_c(\rho)$称为信道的**多普勒功率谱**（Doppler power spectrum）。作为自相关函数的傅里叶变换，它给出了接收信号功率在多普勒频率ρ上的密度。使$|S_c(\rho)|$不为零的最大ρ值称为信道的**多普勒扩展**（Doppler spread），记为B_D。由$A_c(\Delta t)$和$S_c(\rho)$是傅里叶变换关系可得$B_D \approx 1/T_c$。如果发射机和信道中的反射体都是静止的，接收机以速度v移动，那么$B_D \leqslant v/\lambda = f_D$。再注意到窄带衰落模型中抽样在间隔$\Delta t=0.4/f_D$时的抽样独立，因此就一般而言$B_D \approx k/T_c$，$k$的具体取值与$S_c(\rho)$的形状有关。图3-14示出了多普勒功率谱$S_c(\rho)$及其傅里叶反变换$A_c(\Delta t)$。

90

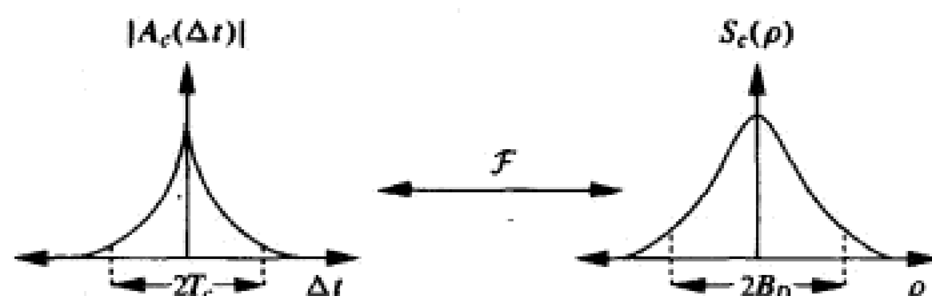


图3-14　多普勒功率谱、多普勒扩展及相干时间

例3.7：某信道的多普勒扩展是$B_D=80\text{Hz}$，求对接收信号进行抽样所得样值近似独立的时间间隔。

解：信道的相干时间为$T_c \approx 1/B_D = 1/80$，因此相隔12.5ms的抽样是近似不相关的。由于信道特性为高斯随机过程，所以这些抽样是近似独立的。

3.3.4　自相关和散射函数的变换

从式（3-61）可以看出，由式（3-53）定义的散射函数$S_c(\tau;\ \rho)$是$S_C(\Delta f;\ \rho)$关于变量

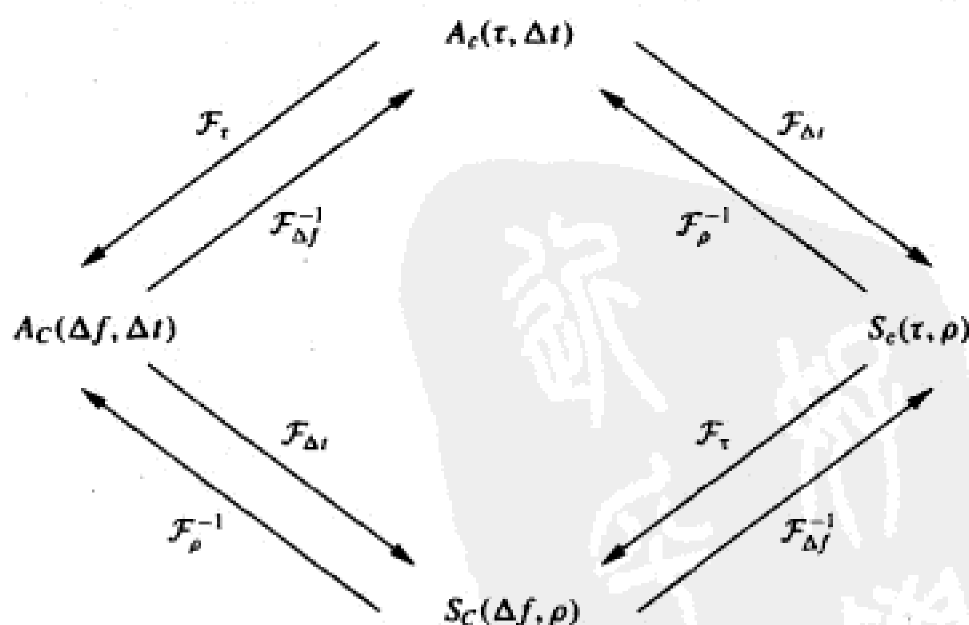


图3-15　傅里叶变换关系

Δf的傅里叶反变换。这样，$S_c(\tau;\rho)$和$A_C(\Delta f;\Delta t)$是一种二重傅里叶变换的关系：

$$S_c(\tau;\rho)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}A_C(\Delta f;\Delta t)\mathrm{e}^{-\mathrm{j}2\pi\rho\Delta t}\mathrm{e}^{\mathrm{j}2\pi\tau\Delta f}\,\mathrm{d}\Delta t\,\mathrm{d}\Delta f \tag{3-63}$$

图3-15显示了4个函数$A_C(\Delta f;\Delta t)$、$A_c(\tau;\Delta t)$、$S_C(\Delta f;\rho)$和$S_c(\tau;\rho)$之间的关系。 91

一般用实测散射函数的方法来近似计算信道的时延扩展、相干带宽、多普勒扩展、相干时间。实测得到经验散射函数$S_c(\tau;\rho)$，再通过傅里叶反变换得到$A_C(\tau;\Delta t)=\mathcal{F}_\rho^{-1}[S_c(\tau;\rho)]$，令$\Delta t=0$得到功率时延谱$A_c(\tau)$，然后计算出平均时延扩展和均方根时延扩展。相干带宽近似为$B_c\approx 1/\sigma_{T_m}$，多普勒扩展近似为使$S_c(0;\rho)$大致不为零的$\rho$的范围，相干时间为$T_c\approx 1/B_D$。

3.4 离散时间模型

对于一些简单的分析，时变冲激响应这样的信道模型往往显得过于复杂。此时可考虑使用宽带多径模型的离散时间近似。由Turin提出[21]的这种离散时间模型特别适合于研究第13章要讲述的扩频系统和RAKE接收机。该模型假设传播环境由一些孤立的散射点组成，如图3-16所示，其多径分量由一些子路径簇构成，同一簇中的分量有近似相同的时延τ_n，它们合为一径，不同簇中的分量有不同的时延$\tau_n\neq\tau_m$，且$|\tau_n-\tau_m|>1/B$，它们是可分辨径，在这里B是信号带宽。式（3-6）的信道模型相应修正为$(N+1)$个路径簇的叠加： 92

$$c(\tau;t)=\sum_{n=0}^{N}\alpha_n(t)\mathrm{e}^{-\mathrm{j}\phi_n(t)}\delta(\tau-\tau_n(t)) \tag{3-64}$$

对于给定的t，接收信号的统计特性由$\{\tau_n\}_0^N$、$\{\alpha_n\}_0^N$和$\{\phi_n\}_0^N$的统计特性确定。可按下面的方法进一步将其简化为离散时间近似模型。对于给定的时间t，把时延轴τ均匀划分为间隔为T的M个时间段，并满足$MT\geqslant\sigma_{T_m}$，σ_{T_m}是信道的均方根时延扩展，可根据经验设定。把各个子路径簇分别归到M个时间段中的某一段中，如图3-17所示。这个离散模型的多径时延扩展为MT，路径分辨能力T由发送信号的带宽决定：$T\approx 1/B$。对于第n个时间段，示性函数$r_n(1\leqslant n\leqslant M)$表示这个时间段中是否存在多径分量：$r_n=1$表示存在，$r_n=0$表示不存在。若$r_n=1$，则相应多径分量的幅度和相位$(\alpha_n,\theta_n)$将服从一定的经验分布。这个经验分布是在传播环境中对各个时间段n在不同位置测量(α_n,θ_n)得到的。对于不同的时间段$n\neq m$，(α_n,θ_n)和(α_m,θ_m)的分布一般不同：有可能是相同类型的衰落，但参数不同，如为不同K因子的莱斯衰落；也可能是不同类型的衰落，如第n个时间段是瑞利衰落，而第m个时间段是Nakagami衰落。

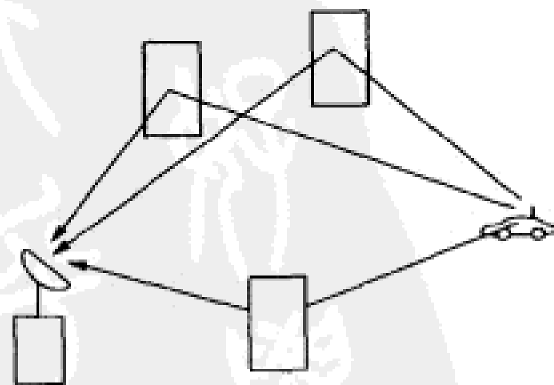


图3-16 散射点信道模型

这样就完成了单个数据快照（即单个时间点）的离散时间统计模型。再用

一系列抓拍的配置数据来体现信道冲激响应的时变性，比如体现接收机以某个速度在城市中移动时所看到的冲激响应的变化。因此，模型不仅要包括每个抓拍（每个时间t内）的(τ_n, α_n, ϕ_n)的一阶统计特性，还必须包括各抓拍之间的时空相关性，模型假设具有马尔科夫性。关于模型细节、N和(τ_n, α_n, ϕ_n)的经验分布，请参见文献[21]。

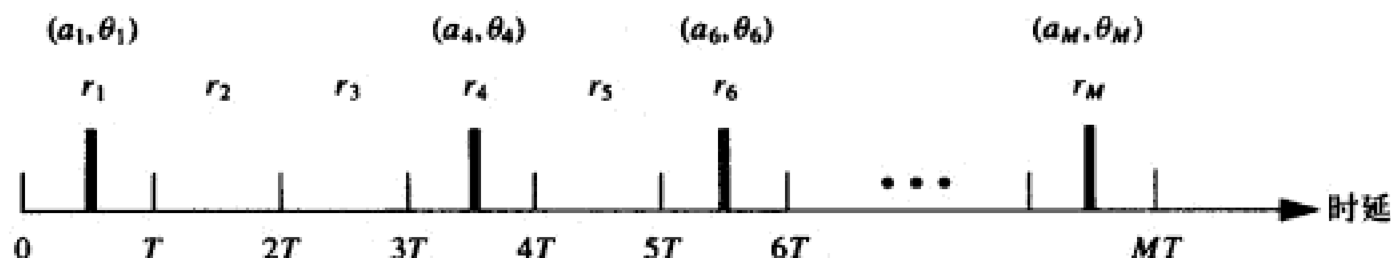


图3-17　离散时间近似

3.5　空时信道模型

由于分集和容量方面的好处，无线系统的发射机和接收机经常配置多个天线。多天线系统要求信道模型必须能同时体现信道的空间（到达角）特性和时间特性。典型的空时模型假设信道由许多产生多径的散射体组成[22~23]，散射体相对于接收机的位置决定多径分量的到达角。空时模型既可以是二维的，也可以是三维的。

93

考虑一个二维的多径传播环境，接收机或发射机配置了M个天线阵元。此时，式(3-6)的时变冲激响应模型可扩展为

$$c(\tau,t)=\sum_{n=0}^{N(t)}\alpha_n(t)\mathrm{e}^{-\mathrm{j}\phi_n(t)}\boldsymbol{a}(\theta_n(t))\delta(\tau-\tau_n(t)) \tag{3-65}$$

从而包括天线阵的到达角。式中$\phi_n(t)$为天线阵原点处的相移、$\boldsymbol{a}(\theta_n(t))$是阵元的响应矢量：

$$\boldsymbol{a}(\theta_n(t))=[\mathrm{e}^{-\mathrm{j}\psi_{n,1}},\ldots,\mathrm{e}^{-\mathrm{j}\psi_{n,M}}]^T \tag{3-66}$$

其中$\psi_{n,i}=[x_i\cos\theta_n(t)+y_i\sin\theta_n(t)]2\pi/\lambda$，($x_i$, y_i)是阵元相对于原点的位置，$\theta_n(t)$为相对于天线阵原点的多径到达角。假定所有径的到达角是平稳同分布，记θ为这个随机的到达角、$A(\theta)$为角度θ处的平均接收功率，则平均角度扩展μ_θ和均方根角度扩展σ_θ可用功率角度谱$A(\theta)$定义为

$$\mu_\theta=\frac{\int_{-\pi}^{\pi}\theta A(\theta)\,\mathrm{d}\theta}{\int_{-\pi}^{\pi}A(\theta)\,\mathrm{d}\theta} \tag{3-67}$$

$$\sigma_\theta=\sqrt{\frac{\int_{-\pi}^{\pi}(\theta-\mu_\theta)^2A(\theta)\,\mathrm{d}\theta}{\int_{-\pi}^{\pi}A(\theta)\,\mathrm{d}\theta}} \tag{3-68}$$

到达角相隔$1/\sigma_\theta$的两个信号近似不相关。更多关于不同传播环境下功率随到达角的分布以及天线阵元间相关性的内容可以参考文献[23]。

从二维模型扩展到三维模型时，多径到达方向同时体现了仰角和方位角。文献[24~26]中提出了一些不同的三维信道模型。文献[22]利用这些思想将空时特性结合到Jakes的均匀散射模型结中。文献[27]中还有其他一些关于空时模型的论文。

习题

3.1　考虑由直射波和地面反射波组成的二径信道，每径上的信号衰减服从自由空间路径损耗模型。发送端是固定的基站，高度为h。接收端固定于一货车上，高度也为h。货车从基站出发，以速度v向远处移动。当收发距离$d=vt$充分大时，反射路径的长度可近似为$r+r'\approx d+2h^2/d$。求接收端相应的时变信道冲激响应。

3.2　写出两径模型中时延扩展T_m的表达式，并给出收发距离充分大时的简化表达式。计算$h_t=10\text{m}$、$h_r=4\text{m}$、$d=100\text{m}$时的T_m值。

3.3　某时不变室内无线信道中直射分量的时延是23ns、第一个多径分量的时延为48ns、第二个多径分量的时延为67ns。针对解调器同步于直射分量或第一个多径分量这样两种情形，分别求相应的时延扩展。

3.4　请针对$f_c=1\text{GHz}$、收发距离大于10m的情形，说明$f_c\tau_n$的最小值远大于1。

3.5　设X和Y是两个相互独立的零均值高斯随机变量，方差同为σ^2，证明：$Z=\sqrt{X^2+Y^2}$服从瑞利分布，Z^2服从指数分布。

94

3.6　假设瑞利衰落信道中的平均信号功率是$2\sigma^2=-80\text{dBm}$，请就门限值分别为$P_0=-95\text{dBm}$和$P_0=-90\text{dBm}$的情形求相应的中断率。

3.7　假设某业务要求门限值为$P_0=-80\text{dBm}$时的中断率为0.01，求瑞利信道下的平均信号功率。

3.8　假设莱斯信道中$2\sigma^2=-80\text{dBm}$，直射信号平均功率为$s^2=-80\text{dBm}$。若目标功率值为$P_0=-80\text{dBm}$，求中断率。

3.9　本题说明莱斯分布和Nakagami分布在尾部有明显的不同。请画出$K=1$、5、10的莱斯分布的累积分布函数以及参数相应为$m=(K+1)^2/(2K+1)$的Nakagami分布的累积分布函数。对于较大的x，莱斯分布或其Nakagami近似是否有较大的中断率$p(\gamma<x)$？

3.10　为了提高蜂窝系统的性能，可让多个基站接收同一移动台的信号，再选择出最强的一个信号，或者以某种优化的权值将这些接收信号叠加在一起。这样做能提高信噪比、减弱阴影的影响。这种技术称为*宏分集*（macrodiversity）。本题考查这种技术的优势。分集的详细内容见第7章。

假设某个移动台位于蜂窝网中两个基站之间的中点处，移动台从基站所收到的信号（dBW）为

$$P_{r,1}=W+Z_1$$
$$P_{r,2}=W+Z_2$$

其中Z_1、Z_2是服从$N(0,\sigma^2)$的随机变量。定义宏分集中的中断为事件$P_{r,1}$和$P_{r,2}$同时小于门限T。

（a）解释$P_{r,1}$、$P_{r,2}$表达式中的W、Z_1、Z_2的含义。

（b）如果Z_1、Z_2独立，证明中断率为$P_{out}=[Q(\Delta/\sigma)]^2$，其中$\Delta=W-T$为移动台位置处的衰落余量。

（c）假设Z_1、Z_2相关，满足：

$$Z_1 = aY_1 + bY$$
$$Z_2 = aY_2 + bY$$

其中Y、Y_1、Y_2是相互独立的随机变量，服从$N(0, \sigma^2)$，且a、b满足$a^2+b^2=1$，证明

$$P_{\text{out}} = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \left[Q\left(\frac{\Delta + by\sigma}{|a|\sigma} \right) \right]^2 e^{-y^2/2} \, dy$$

（d）假设$a=b=1/\sqrt{2}$、$\sigma=8$、$\Delta=5$，比较（b）和（c）中的中断率。（需要进行数值

95

积分）

3.11　本题的目的是建立移动通信信道中的瑞利衰落仿真器，方法是让高斯过程通过一个滤波器，该滤波器根据3.2.1节中所给出的同相分量和正交分量的功率谱密度设计。请按以下步骤做：

（a）编写仿真代码，用以产生幅度随时间瑞利衰落的信号。要求：取样率至少为每秒1000样值，包络的平均值为1，仿真参数包括多普勒频率f_D。最容易的方法是用Matlab，其他代码也可。

（b）写出仿真说明文档，说明衰落包络产生的详细过程，要求有流程图及必要的公式。

（c）提交调好的代码。

（d）提交$f_D=1$、10、100Hz时，两秒钟时间内，接收幅度（dB）随时间变化的图。

3.12　设瑞利衰落信道的平均功率为$\overline{P}_r=30$dB，多普勒频率为$f_D=10$Hz，计算当目标衰落值分别为$P_0=0$dB、$P_0=15$dB及$P_0=30$dB时的平均衰落时长。

3.13　设瑞利衰落的平均功率为$\overline{P}_r$，请推导出接收功率高于目标衰落值P_0的平均持续时间。求出$\overline{P}_r=20$dB、$P_0=25$dB、$f_D=50$Hz时的平均持续时间数值。

3.14　设瑞利衰落信道的平均功率为$\overline{P}_r=10$dB，多普勒频率为$f_D=80$Hz。考虑用八状态的有限状态马氏模型来近似这个信道，假设马氏模型中的时间间隔为$T=10$ms。八个区间取为：$R_1=\{\gamma: -\infty<\gamma<-10\text{dB}\}$、$R_2=\{\gamma: -10\text{dB}\leqslant\gamma<0\text{dB}\}$、$R_3=\{\gamma: 0<\gamma<5\text{dB}\}$、$R_4=\{\gamma: 5\text{dB}\leqslant\gamma<10\text{dB}\}$、$R_5=\{\gamma: 10\text{dB}\leqslant\gamma<15\text{dB}\}$、$R_6=\{\gamma: 15\text{dB}\leqslant\gamma<20\text{dB}\}$、$R_7=\{\gamma: 20\text{dB}\leqslant\gamma<30\text{dB}\}$、$R_8=\{\gamma: 30\text{dB}\leqslant\gamma<\infty\}$。求马氏模型的状态转移概率。

3.15　通过发送900MHz正弦波测得某信道的散射函数为：

$$S(\tau,\rho) = \begin{cases} \alpha_1\delta(\tau) & \rho = 70\text{Hz} \\ \alpha_2\delta(\tau - 0.022\mu\text{s}) & \rho = 49.5\text{Hz} \\ 0 & \text{其他} \end{cases}$$

其中α_1，α_2取决于路径损耗、阴影效应和多径衰落。这个散射函数明显对应两径模型。设发射机和接收机的高度都是8m。

（a）求发射机和接收机之间的距离和相对速度。

（b）对于（a）中算出的距离，路径损耗是随d^2还是d^4下降？（提示：注意是两径信道。）

（c）带宽为30kHz的话音调制信号经过此信道传输时，所经历的是平衰落还是频率选择

96

性衰落？

3.16 一宽带信道的自相关函数为：

$$A_c(\tau, \Delta t) = \begin{cases} \text{sinc}(W\Delta t) & 0 \leqslant \tau \leqslant 10\,\mu\text{s} \\ 0 & \text{其他} \end{cases}$$

其中$W=100\text{Hz}$，$\text{sinc}(x)=\sin(\pi x)/(\pi x)$。

(a) 这个信道是室内信道还是室外信道？为什么？

(b) 画出该信道的散射函数。

(c) 计算信道的平均时延扩展、均方根时延扩展和多普勒扩展。

(d) 当数据速率近似超过什么范围时，信号在此信道传送会发生频率选择性衰落？

(e) 此信道是瑞利衰落还是莱斯衰落？为什么？

(f) 假设是瑞利衰落，那么信号功率持续低于平均值的平均时长是多少？

(g) 假设信道中传输的是窄带二进制调制信号。系统采用了某种纠错编码，它能纠正两比特错误。假设接收功率在低于平均值时，一定会发生错误；而当功率等于或者大于平均值时，一定不会发生错误。若信道为瑞利衰落，求无误码传输的最大数据速率。可近似认为最大衰落时长不超过平均衰落时长的两倍。

3.17 设散射函数$S_c(\tau,\rho)$在$0\leqslant\tau\leqslant0.1\text{ms}$及$-0.1\leqslant\rho\leqslant0.1\text{Hz}$范围内不为零，且在此范围内近似为常数。

(a) 求此信道的多径时延扩展和多普勒扩展。

(b) 假设信道输入是频差为Δf的两个正弦波[1]，求能使这两个正弦波的信道响应近似独立的最小Δf值。

(c) 设信道输入为两个正弦波：$u_1(t)=\sin 2\pi ft$、$u_2(t)=\sin 2\pi ft(t+\Delta t)$，求能使$u_1(t)$和$u_2(t)$的输出响应近似独立的最小$\Delta t$值。

(d) 对于典型的3kHz带宽语音信道，该信道呈现为平衰落还是频率选择性衰落？如果是蜂窝网中带宽为30kHz的信道又如何？

参考文献

[1] T. S. Rappaport, *Wireless Communications – Principles and Practice*, 2nd ed., Prentice-Hall, Englewood Cliffs, NJ, 2001.

[2] D. Parsons, *The Mobile Radio Propagation Channel*, Wiley, New York, 1994.

[3] R. S. Kennedy, *Fading Dispersive Communication Channels*, Wiley, New York, 1969.

[4] R. H. Clarke, "A statistical theory of mobile radio reception," *Bell System Tech. J.*, pp. 957–1000, July/August 1968.

[5] W. C. Jakes, Jr., *Microwave Mobile Communications*, Wiley, New York, 1974.

[6] M. K. Simon and M.-S. Alouini, *Digital Communication over Fading Channels: A Unified Approach to Performance Analysis*, Wiley, New York, 2000.

[7] G. L. Stuber, *Principles of Mobile Communications*, 2nd ed., Kluwer, Dordrecht, 2001.

[8] M. Pätzold, *Mobile Fading Channels*, Wiley, New York, 2002.

[9] S. O. Rice, "Mathematical analysis of random noise," *Bell System Tech. J.*, pp. 282–333, July 1944, and pp. 46–156, January 1945.

[10] F. Babich, G. Lombardi, and E. Valentinuzzi, "Variable order Markov modeling for LEO mobile satellite channels," *Elec. Lett.*, pp. 621–3, April 1999.

97

1. 原书错误，此处已按勘误表改正。——译者注

[11] A. M. Chen and R. R. Rao, "On tractable wireless channel models," *Proc. Internat. Sympos. Pers., Indoor, Mobile Radio Commun.*, pp. 825–30, September 1998.

[12] H. S. Wang and N. Moayeri, "Finite-state Markov channel – A useful model for radio communication channels," *IEEE Trans. Veh. Tech.*, pp. 163–71, February 1995.

[13] C. C. Tan and N. C. Beaulieu, "On first-order Markov modeling for the Rayleigh fading channel," *IEEE Trans. Commun.*, pp. 2032–40, December 2000.

[14] C. Pimentel and I. F. Blake, "Modeling burst channels using partitioned Fritchman's Markov models," *IEEE Trans. Veh. Tech.*, pp. 885–99, August 1998.

[15] Y. L. Guan and L. F. Turner, "Generalised FSMC model for radio channels with correlated fading," *IEE Proc. Commun.*, pp. 133–7, April 1999.

[16] M. Chu and W. Stark, "Effect of mobile velocity on communications in fading channels," *IEEE Trans. Veh. Tech.*, pp. 202–10, January 2000.

[17] C. Komninakis and R. D. Wesel, "Pilot-aided joint data and channel estimation in flat correlated fading," *Proc. IEEE Globecom Conf.*, pp. 2534–9, November 1999.

[18] M. Peleg, S. Shamai (Shitz), and S. Galan, "Iterative decoding for coded noncoherent MPSK communications over phase-noisy AWGN channels," *IEE Proc. Commun.*, pp. 87–95, April 2000.

[19] P. A. Bello, "Characterization of randomly time-variant linear channels," *IEEE Trans. Commun. Syst.*, pp. 360–93, December 1963.

[20] W. C. Y. Lee, *Mobile Cellular Telecommunications Systems*, McGraw-Hill, New York, 1989.

[21] G. L. Turin, "Introduction to spread spectrum antimultipath techniques and their application to urban digital radio," *Proc. IEEE*, pp. 328–53, March 1980.

[22] Y. Mohasseb and M. P. Fitz, "A 3-D spatio-temporal simulation model for wireless channels," *IEEE J. Sel. Areas Commun.*, pp. 1193–1203, August 2002.

[23] R. Ertel, P. Cardieri, K. W. Sowerby, T. Rappaport, and J. H. Reed, "Overview of spatial channel models for antenna array communication systems," *IEEE Pers. Commun. Mag.*, pp. 10–22, February 1998.

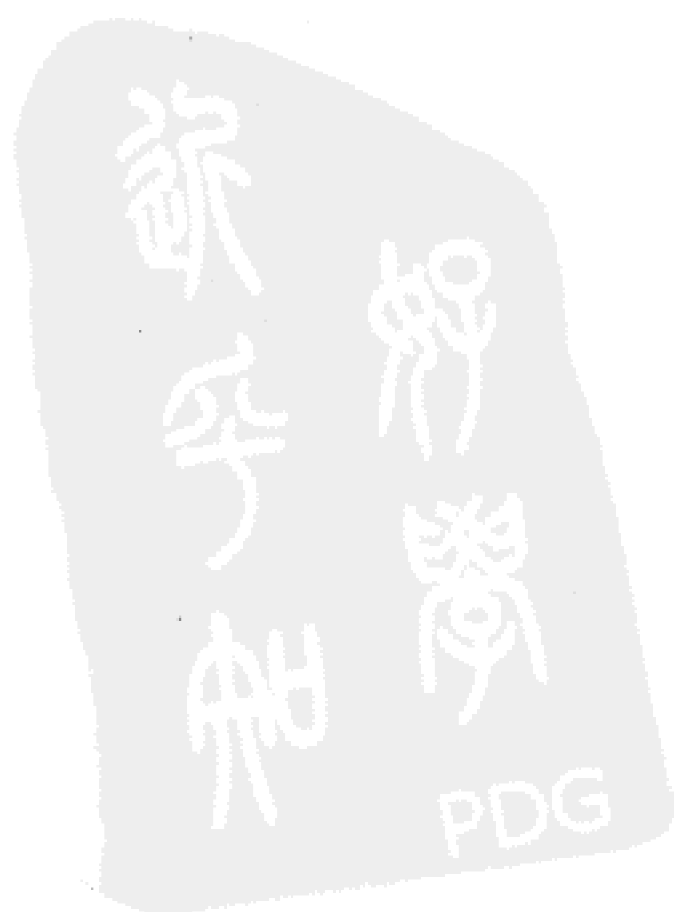
[24] T. Aulin, "A modified model for fading signal at the mobile radio channel," *IEEE Trans. Veh. Tech.*, pp. 182–202, August 1979.

[25] J. D. Parsons and M. D. Turkmani, "Characterization of mobile radio signals: Model description," *Proc. IEE*, pt. 1, pp. 549–56, December 1991.

[26] J. D. Parsons and M. D. Turkmani, "Characterization of mobile radio signals: Base station cross-correlation," *Proc. IEE*, pt. 1, pp. 557–65, December 1991.

98

[27] L. G. Greenstein, J. B. Andersen, H. L. Bertoni, S. Kozono, and D. G. Michelson, Eds., *IEEE J. Sel. Areas Commun.*, Special Issue on Channel and Propagation Modeling for Wireless Systems Design, August 2002.



第4章 无线信道的信道容量

无线通信不断增长的需求使研究各种无线系统潜在的信道容量极限变得越来越重要。信道容量给出了不考虑编译码时延和复杂度的条件下，使误码率趋于无限小的最高的传输速率。信道容量的数学理论最早是由香农（Claude Shannon）于20世纪40年代末建立的。该理论的基础概念是信道输入和信道输出之间的互信息量[1~3]。香农将信道容量定义为这个互信息量在不同的信道输入分布下的最大值。这些数学概念的重要性在于香农的编码定理及其逆定理。编码定理证明，一定存在这样一种编码方案，其数据传输速率接近信道容量而误码率可忽略；编码定理的逆定理证明，对于任何编码方案，若传输速率大于信道容量，则误码率必不可能接近于零。香农的许多想法对当时来说是非常有革命性的：他预言了电话信道中的高数据速率，他还指出不用降低速率或增加带宽就能降低误码率。今天先进的调制编码技术已经证实了香农的理论。在电话线上，我们已经能以接近香农容量的数据速率传输而保持非常低的误码率。本书第5章和第8章将分别讨论这些先进的调制和编码技术。

本章我们将讨论单用户单天线系统（发射机和接收机都只有一副天线）的信道容量。单用户多天线以及多用户系统的容量分别在第10章和第14章介绍。本章的讨论将包括时变信道和时不变信道的情况。我们首先介绍大家都很熟悉的时不变加性高斯白噪声（additive white Gaussian noise，AWGN）信道的容量公式，然后研究时变平衰落信道的容量。与AWGN信道不同的是，平衰落信道的容量并不是由单个公式给出的，因为信道容量还依赖于接收端和发送端对时变信道信息的了解。此外不同的信道信息假设下的信道容量的定义也不同，容量所描述的可以是所有衰落状态下平均速率的最大值，也可以是所有衰落状态下（或者以某个中断率）都能达到的固定速率的最大值。

本章首先考虑接收端和发送端只知道衰落分布时平衰落信道的容量。除个别特例外，这一假设下的信道容量一般很难求解。接下来我们考虑信道衰落值仅对于接收端已知（通过接收端的信道估计）的情形和对于收发端均已知（通过接收端信道估计及发射机的反馈）的情形。我们将看到，当收发两端都已知信道衰落信息时，达到衰落信道容量的方式是根据信道的变化自适应地调整发送功率、速率及编码方式，其中最佳的功率分配是时域注水法（water-filling）功控，它在信道条件好的时候增大功率及数据速率，信道条件差的时候减小功率及数据速率。

本章还将考虑频率选择性衰落信道的容量。时不变频率选择性信道的容量存在解，达到容量的方式是频域的而非时域的注水法功控。时变频率选择性信道的容量一般难以求解，不过可将其近似为一组独立的平衰落子信道，子信道间进行最优功率分配，则所求信道容量是这些子信道容量之和。此时最佳的功率分配是时频二维的注水法功控。

本章只考虑离散时间系统。大多数连续时间的系统都可以通过采样转变为离散时间系统，并保持容量不变。但注意对于时变信道，为了保持容量不变，时变性可能会要求提高采样率[4]，所以在选取合适抽样率时一定要慎重。

4.1 AWGN信道容量

考虑一个离散时间加性高斯白噪声（additive white Gaussian noise，AWGN）信道，信道的输入输出关系是$y[i]=x[i]+n[i]$，其中$x[i]$是信道在时刻i的输入，$y[i]$是相应的输出，随机过程$n[i]$是高斯白噪声。假设信道带宽为B，接收信号功率是P。接收信噪比（Signal-to-noise ratio，SNR）定义为$x[i]$的功率除以$n[i]$的功率，是恒定值$\gamma=P/(N_0B)$。白噪声的单边功率谱密度是$N_0/2$。AWGN信道的容量由著名的香农公式给出[1]：

$$C = B\log_2(1+\gamma) \tag{4-1}$$

容量的单位是比特/秒（bit/s）。香农编码定理证明，存在这样一种编码方案，其数据传输速率任意接近信道容量，而误比特率任意小。编码定理的逆定理指出，若数据传输速率$R>C$，则无论何种编码都不可能使误比特率任意小。这两个定理的证明用到了信道输入和输出之间的互信息量。考虑一个离散无记忆时不变信道，若信道的随机输入是x，随机输出是y，则互信息量定义为

100

$$I(X;Y) = \sum_{x\in\mathcal{X},y\in\mathcal{Y}} p(x,y)\log\left(\frac{p(x,y)}{p(x)p(y)}\right) \tag{4-2}$$

其中的求和对所有可能的输入$x\in\mathcal{X}$和输出$y\in\mathcal{Y}$进行，$\mathcal{X}$和$\mathcal{Y}$是x和y的取值集合。式中的log通常以2为底，这样互信息量的单位就是比特/秒。互信息量也可以用输出y和条件输出$y|x$的熵来表示：$I(X;Y)=H(Y)-H(Y|X)$，其中$H(Y)=-\sum_{y\in\mathcal{Y}}p(y)\log p(y)$，$H(Y|X)=-\sum_{x\in\mathcal{X},y\in\mathcal{Y}}p(x,y)\log p(y\mid x)$。香农证明，信道容量等于遍历所有可能的输入分布后得到的最大互信息量：

$$C = \max_{p(x)} I(X;Y) = \max_{p(x)} \sum_{x,y} p(x,y)\log\left(\frac{p(x,y)}{p(x)p(y)}\right) \tag{4-3}$$

对于AWGN信道，式（4-3）中的求和变为积分，使容量最大的输入分布是高斯分布，这时的信道容量由式（4-1）给出。对有记忆信道，互信息量和信道容量由输入输出序列x^n和y^n决定。有关信道容量、互信息量、编码定理及其逆定理的详细介绍可参见文献[2, 5~6]。

信道编码定理及其逆定理的理论证明并没有考虑通信系统的复杂性及时延问题。考虑到实际通信系统中存在的一些限制，香农容量可作为数据速率的上界。在香农提出信息论的时候，普通电话线上可传输的数据速率为100bit/s左右，香农按照普通电话线的条件计算出的信道容量是30kbit/s。因此当时的人们认为香农的理论对实际系统来说，是一个没有什么用处的上界。随着硬件、调制、编码等技术领域的突破性进展，今天的调制解调器速率已经可以非常接近香农于20世纪50年代预言的数值。实际上，在某些电话线上调制解调器的速率甚至超过了30kbit/s，这是因为现在的传输线路质量要比香农时代的好，因而接收功率也比香农当初用于计算的接收功率大。此外，Turbo码在AWGN信道中能达到和香农极限仅差零点几个分贝的性能[7]。

无线信道一般表现为平衰落或频率选择性衰落。我们将在下面两节中考虑这两种衰落信道在不同假设条件下的信道容量。

例4.1：假设某无线信道中功率随距离下降的关系为$P_r(d)=P_t(d_0/d)^3$，其中$d_0=10$m。设信道带宽为$B=30$kHz，AWGN的功率谱密度为$N_0/2$，其中$N_0=10^{-9}$W/Hz。设发送功率为1W，求接收端与发送端相距100m和1km时的信道容量。

解：$d=100$m时接收信噪比为$\gamma=P(d)/(N_0B)=0.1^3/(10^{-9}\times30\times10^3)\approx33\approx15$dB，代入$C=B\log_2(1+\gamma)$得容量为$C=3000\log_2(1+33)=152.6$kbit/s；$d=1$km时$\gamma=P(d)/(N_0B)=0.01^3/(10^{-9}\times30\times10^3)\approx0.033\approx-15$dB，容量为$C=30\ 000\log_2(1+0.033)=1.4$kbit/s。注意由于路径损耗指数是3，距离的增加大大降低了接收信号的功率，从而大大降低了容量。

101

4.2 平坦衰落信道的容量

4.2.1 信道和系统模型

考虑如图4-1所示的平稳遍历离散时间信道，时变的信道增益为$\sqrt{g[i]}$，$g[i]\geqslant0$，$n[i]$是加性高斯白噪声。信道的功率增益$g[i]$服从某种分布$p(g)$，例如对于瑞利衰落其增益服从指数分布。假设$g[i]$与信道的输入无关。$g[i]$能随时间i变化，$g[i]$可能是一个独立同分布序列，也可能是一个相关序列。在分块衰落信道（block fading channel）中，$g[i]$在某一分块时间T内保持不变，之后依其分布$p(g)$变为另一个独立的值。令$\overline{P}$表示平均发送功率，$N_0/2$表示AWGN的功率谱密度，B代表接收信号带宽，则瞬时接收信噪比为$\gamma[i]=\overline{P}g[i]/(N_0B),0\leqslant\gamma[i]<\infty$，其均值为$\overline{\gamma}=\overline{P}\overline{g}/(N_0B)$。由于$\overline{P}/(N_0B)$恒定，$g[i]$的分布就决定了$\gamma[i]$的分布，反之亦然。

如图4-1所示，输入消息w通过编码器编码为码字x，x的一个符号$x[i]$通过时变信道在i时刻传输。信道增益$g[i]$，也称为信道边信息（channel side information, CSI），在码字的传输过程中随时间发生变化。$\hat{w}$是接收端根据接收信号译出的消息。

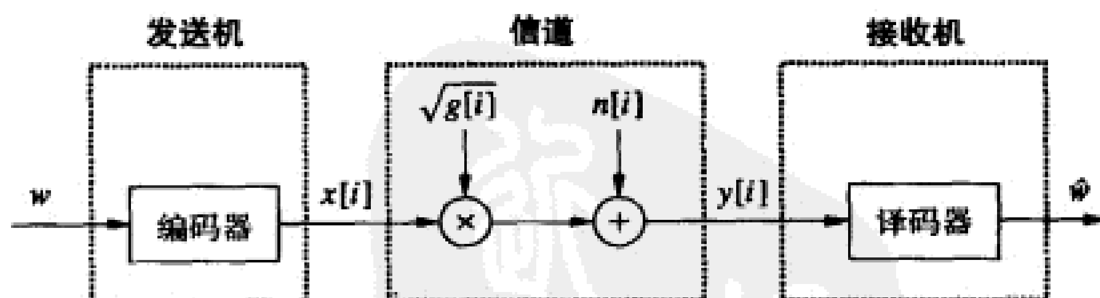


图4-1 平衰落信道及系统模型

信道容量取决于发送端和接收端对$g[i]$的了解情况，分以下三种：

（1）信道分布信息（CDI）已知：发送端和接收端都已知$g[i]$的分布。

（2）接收端已知CSI：接收端已知$g[i]$在时刻i的值，且发送端和接收端都已知$g[i]$的分布。

（3）发送端和接收端都已知CSI：发送端和接收端都已知$g[i]$在时刻i的值及$g[i]$的分布。

发送端和接收端都已知CSI时，发送端可根据信道i时刻的增益实时调整其发送功率和传输速率，从而能达到最大的信道容量。注意瞬时信噪比$\gamma[i]$为$g[i]$和常数$\overline{P}/(N_0B)$的乘积，所以知道$g[i]$的CSI或CDI等于知道$\gamma[i]$的CSI或CDI。以上三种情况之外的时变信道容量问题请参考文献[8~9]。

4.2.2 信道分布信息已知

首先考虑信道增益的分布$p(g)$，也即信噪比的分布$p(\gamma)$对于收发两端都已知的情形。若信道增益为独立同分布，则信道容量由式（4-3）给出，但由于衰落分布的复杂性，很难求解出使式（4-3）达到最大的输入分布$p(x)$。此外，相关衰落会引入信道记忆，此时需要按输入序列而不是按符号来求最佳的输入分布，这将使求解更为困难。由于上述原因，除个别特殊的信道以外，信道分布信息已知情况下找出能使衰落信道容量最大的输入分布问题仍然是一个尚未解决的问题。

已知CDI时，可到达容量的输入分布问题在两种情形下是可以求解的，一是独立同分布的瑞利衰落信道，二是*有限状态马尔科夫*（FSMC）信道。在独立同分布的瑞利衰落信道中，功率增益服从指数分布，并且对于不同的信道使用，信道增益是独立变化的。此时，文献[10]指出最优分布的信道输入是一个有限取值的离散随机变量，取值包括0。具体分布函数及对应的信道容量并无闭式解，只能以数值方法求得。对于遵循如此常见的衰落分布，而且没有任何相关性因素的情况，最优分布及相应的容量居然没有闭式解，这多少有些出人意料。对于信道分布不一定是瑞利或不一定独立同分布的情形，文献[11]给出了容量的上界和下界，这些界在高信噪比下比较准确。

3.2.4节中曾经提到，FSMC模型是对相关瑞利衰落模型的一种马尔科夫近似。衰落的马尔科夫性表明当前的衰落只取决于前一时刻的衰落，但为了达到信道容量，最佳译码器必须要根据当前的信道输出和过去所有的信道输出来联合译码，这使容量分析大为复杂。文献[12~13]推导了独立同分布的信道输入下FSMC信道的容量，文献[14]给出了任意信道输入分布下的容量。FSMC的容量取决于信道在所有过去输入和输出条件下的极限分布，可用迭代的方法求出。如同独立同分布瑞利衰落的情形，对于FSMC这样简单的分布，信道容量的分析也非常复杂。由此可见，仅CDI已知时，信道容量的分析以及由此寻求相应的设计思路都是非常困难的。

4.2.3 接收端已知CSI

我们现在考虑接收端已知$g[i]$，等同于已知$\gamma[i]$的情形，并假设发送端和接收端都已知$g[i]$的分布。此时有两种信道容量的定义：一种是香农容量，也叫*遍历容量*（ergodic capacity），另一种是*带中断的容量*（capacity with outage），这两种容量对实际系统设计都有重要意义。如同AWGN信道，香农容量定义为可使误码率任意小的最大数据传输速率。注意香农容量所给出的是一个恒定速率，发送端的发送方式不随CSI的变化而调整。由于发送方式没有考虑到具体的信道状态，所以差的信道状态将使香农容量降低。接收端已知CSI时，衰落信道容量的另一种定义是带中断的容量，即指定中断率下信道能传送的最大的恒定传输速率，这里的中断指接收端译码的错误率不是无限小。带中断

的容量的基本前提是，当除信道处于慢深衰落以外，信道都能传送高数据率并获得正确译码。带中断的容量表明了这样一个意思，即若无视个别深衰落状态，不在乎这些时候丢失数据的话，我们就可以发送一个更高的固定速率。带中断的容量含义中并不要求所有瞬时都能正确译码，它存在数据丢失，这一点和香农容量不同，后者所发送的数据要求没有任何丢失。中断率反映了数据的丢失率，也即深衰落的概率。

1. 香农容量（遍历容量）

对于给定平均功率限制$\overline{P}$且接收端已知CSI的衰落信道，文献[15]给出香农容量为

$$C=\int_0^{\infty} B\log_2(1+\gamma)p(\gamma)\,\mathrm{d}\gamma \tag{4-4}$$

注意此式是统计平均：信噪比为γ的AWGN信道的容量是$B\log_2(1+\gamma)$，再按信噪比的分布$p(\gamma)$求平均就得到C，这也是称其为遍历容量的原因。特别注意不能把式（4-4）理解成：当瞬时信噪比为γ时，发送速率设为$B\log_2(1+\gamma)$，最后平均下来就达到了香农容量。因为发送端不知道瞬时γ，信道的传输速率只能是恒定的，与瞬时的γ无关。还要注意，达到此容量的编码必须足够长，使得接收到的码字能遍历衰落信道的所有状态。这会导致显著的译码时延。

由Jensen不等式

$$\begin{aligned}E[B\log_2(1+\gamma)]&=\int B\log_2(1+\gamma)p(\gamma)\,\mathrm{d}\gamma\\&\leqslant B\log_2(1+E[\gamma])=B\log_2(1+\bar{\gamma})\end{aligned} \tag{4-5}$$

其中$\overline{\gamma}$是信道的平均信噪比，可见对于相同的平均信噪比，接收端已知CSI时的香农容量要比AWGN信道容量小，即仅接收端已知CSI时衰落将使容量减小。此外，发送端不知道CSI时，编码设计必须要考虑到信道的相关性，而最大似然译码的复杂度和信道的去相关的时间成正比。另外，如果接收端的CSI不精确，容量将显著下降[16]。

例4.2：某平衰落信道的增益$\sqrt{g[i]}$为独立同分布，且$\sqrt{g[i]}$有三个可能取值：以概率$p_1=0.1$取$\sqrt{g_1}=0.05$、以概率为$p_2=0.5$取$\sqrt{g_2}=0.5$、以概率$p_2=0.4$取$\sqrt{g_3}=1$。发送功率为10mW，噪声的功率谱密度为$N_0/2$，$N_0=10^{-9}$W/Hz，信道带宽为30kHz。假设接接收端知道各个瞬时的$g[i]$，但发送端未知。求该信道的香农容量并与相同平均信噪比的AWGN信道容量作比较。

104

解：接收端有三种可能的信噪比：$\gamma_1=P_t g_1/(N_0 B)=0.8333$、$\gamma_2=P_t g_2/(N_0 B)=83.33$及$\gamma_3=P_t g_3/(N_0 B)=333.33$。对应的出现概率分别为$p_1=0.1$、$p_2=0.5$、$p_3=0.4$，因此香农容量为

$$\begin{aligned}C=\sum_i B\log_2(1+\gamma_i)p(\gamma_i)&=30\,000[0.1\times\log_2(1.8333)+0.5\log_2(84.333)+0.4\log_2(334.33)]\\&=199.26\text{kbit/s}\end{aligned}$$

信道的平均信噪比为$\overline{\gamma}=0.1\times0.8333+0.5\times83.33+0.4\times333.33=175.08=22.43$dB，对应的AWGN信道的容量为$C=B\log_2(1+175.08)=223.8$kbit/s，比相同平均信噪比下接收端已知CSI的平衰落信道容量约大25kbit/s。

2. 带中断的容量

带中断的容量适用于慢变信道，信道的瞬时信噪比在一段传输突发的时间内是恒定

的，经过这段时间后依相应的衰落分布变为另一个值。在这种信道中，如果发送端已知此信道突发时间内的信噪比为γ，就能以速率$B\log_2(1+\gamma)$传输数据而误码率可以做到任意小[1]。由于发送端不知道这个信噪比的数值，所以只能以一个不依赖于瞬时信噪比数值的固定速率进行传输。

带中断的容量允许在某个突发时段以一定概率译错所传输的比特。发送端确定一个最小接收信噪比γ_{min}，再按这个信噪比确定一个速率$C=B\log_2(1+\gamma_{min})$，然后在所有突发中以这个速率传输。如果接收的瞬时信噪比大于或等于γ_{min}，则能正确译码[17~18]。若接收信噪比小于γ_{min}，就不能以接近1的概率译对突发中的所有数据比特，此时接收机将指示出现了一次中断。出现中断的概率为$P_{out}=P(\gamma<\gamma_{min})$。在所有突发中，正确传输的概率是$1-P_{out}$，所以平均正确接收的数据速率为$C_{out}=(1-P_{out})B\log_2(1+\gamma_{min})$。数值$\gamma_{min}$是一个设计参数，取决于可接受的中断率。一般把带中断的容量画成容量随中断率变化的曲线，如图4-2所示。图中针对平均信噪比为$\bar{\gamma}=20$dB的瑞利衰落信道（γ服从指数分布），绘出了归一化容量$C/B=\log_2(1+\gamma_{min})$随中断率$P_{out}=P(\gamma<\gamma_{min})$变化的曲线。可以注意到，当中断率很小时，容量趋于零，这是因为很小的中断率要求在深衰落情况下也要保证正确译码。随着中断率的升高，容量显著增加。不过高中断率下的高速率对应着更高的错误接收率。可以通过优化γ_{min}或P_{out}来最大化平均正确接收速率。

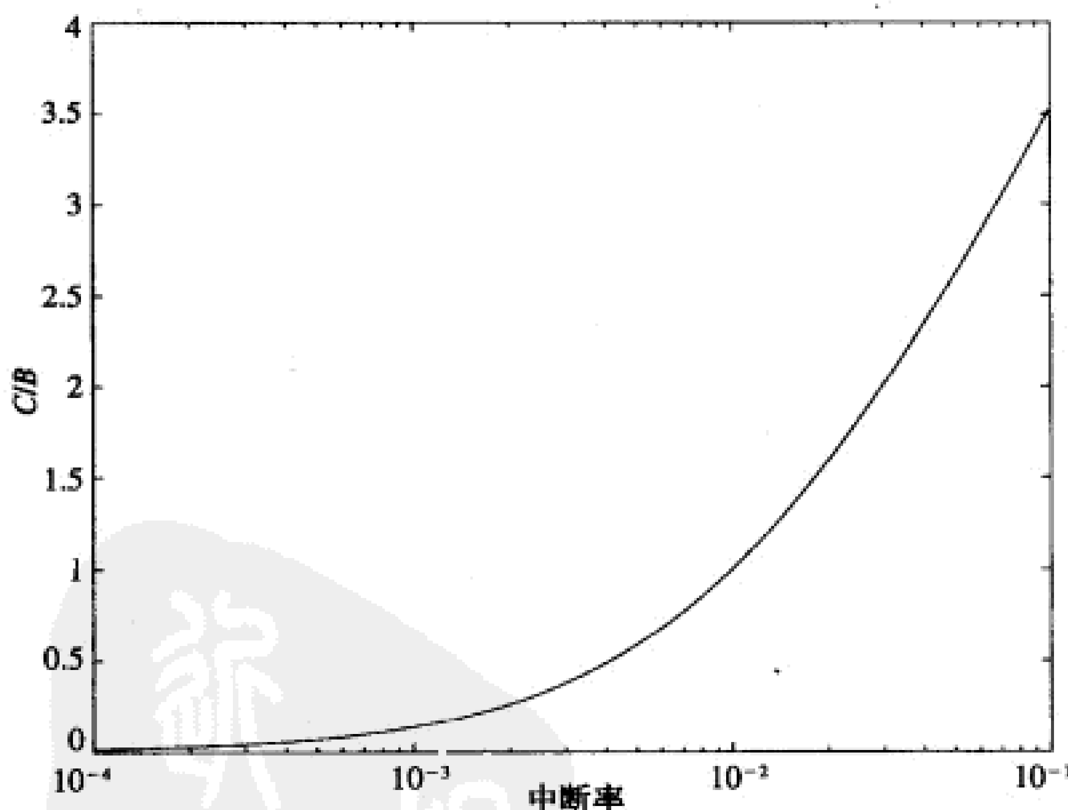


图4-2 归一化容量（C/B）与中断率的关系

例4.3：考虑与前例相同的信道，带宽为30kHz，有三个可能的接收信噪比：出现$\gamma_1=0.8333$的概率是$p_1=0.1$、出现$\gamma_2=83.33$的概率是$p_2=0.5$、出现$\gamma_3=333.33$的概率是$p_3=0.4$。求中断率分别为$P_{out}<0.1$、$P_{out}=0.1$及$P_{out}=0.6$时带中断的容量及对应的平均正确接收速率。

解：因为该时变信道的信噪比是离散的，所以带中断的容量是中断率的阶梯函数。

1. 需假设信噪比固定的这个突发时间足够长，因为能达到容量的编码必须有足够的码长。

P_{out}<0.1表示在所有可能的信道状态下都能正确译码。满足这一点的最小信噪比是$\gamma_{min}=\gamma_1$，对应的容量为$C=B\log_2(1+\gamma_{min})=26.23$kbit/s。当中断率在$0.1\leq P_{out}<0.6$范围内时，信道处在最坏的$\gamma_1$状态时不能正确译码，此时$\gamma_{min}=\gamma_2$，对应的容量为$C=B\log_2(1+\gamma_{min})=191.94$kbit/s。当中断率在$0.6\leq P_{out}<1$范围内时，若接收端信噪比为$\gamma_1$或$\gamma_2$时将不能译码，此时$\gamma_{min}=\gamma_3$，对应的容量为$C=B\log_2(1+\gamma_{min})=251.55$kbit/s。

$P_{out}<0.1$时的平均正确接收速率接近26.23kbit/s，因为此时中断率实际上是0。对于$P_{out}=0.1$，传输速率接近$C=191.94$kbit/s，只有当信噪比为γ_2或γ_3时才能正确译码，所以平均正确接收速率为（1－0.1）×191.94＝172.75kbit/s。对于$P_{out}=0.6$，传输速率接近$C=251.55$kbit/s，只有当信噪比为γ_3时才能正确译码，所以平均正确接收速率为（1－0.6）×251.55＝100.62kbit/s。计算结果表明，对于这个信道比较理想的工程设计是设计传输速率为接近191.94kbit/s，此时出错概率最多只有10%，而数据速率要比无中断的情形高出接近一个数量级。不过对于有些应用，10%的出错重传概率仍显得太大，此时应将发送速率设计为26.23kbit/s且不做重传。关于可接受重传率的设计问题将在第14章中讨论。

106

4.2.4 收发两端都已知CSI

当收发两端都知道CSI时，发送端可根据CSI实时调整发送策略，如图4-3所示。这种情况下不存在带中断的容量，因为发送端已知CSI后，它将不会以接收端会译错的过高速率发送数据。本节中我们将推导根据CSI进行最佳功控及速率自适应时的香农容量，还将介绍其他一些容量定义及相应的功率和速率自适应策略。

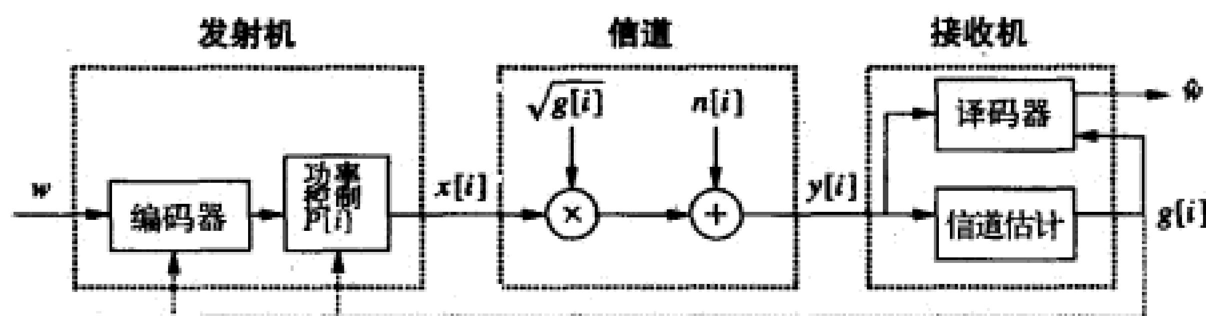


图4-3 收发两端都已知CSI时的系统模型

1. 香农容量

现在我们来讨论i时刻接收端及发送端均知道信道功率增益$g[i]$条件下的香农容量。Wolfowitz最早研究了收发两端都已知CSI的情况下，时变信道的香农容量。按照Wolfowitz的模型，平稳遍历随机过程$s[i]$表示离散无记忆信道在i时刻的信道状态，$s[i]$取值于离散无记忆信道的有限集合S。令C_s代表信道状态为$s\in S$时的容量，$p(s)$代表s出现概率，也即状态s出现的时间比率。则时变信道的容量由文献[19]的定理4.6.1给出为

$$C=\sum_{s\in S}C_s p(s) \tag{4-6}$$

现在我们将上式应用到图4-3所示的系统模型中。信噪比为γ的AWGN信道的容量是$C_\gamma=B\log_2(1+\gamma)$。令$p(\gamma)=p(\gamma[i]=\gamma)$代表接收信噪比的分布。由式（4-6）得收发都已知

CSI时衰落信道的容量为[1]

$$C=\int_0^\infty C_\gamma p(\gamma)\,\mathrm{d}\gamma=\int_0^\infty B\log_2(1+\gamma)p(\gamma)\,\mathrm{d}\gamma \tag{4-7}$$

可见，如果不进行功率控制的话，式（4-4）与式（4-7）是一样的。这表明，如果发送107端不进行功率控制，则让发送端已知CSI并不能增加信道容量。

现在我们允许瞬时的发送功率$P(\gamma)$随γ变化，并受限于平均功率$\bar{P}$：

$$\int_0^\infty P(\gamma)p(\gamma)\,\mathrm{d}\gamma\leqslant\bar{P} \tag{4-8}$$

仅由这个约束条件还不能直接从式（4-7）得到容量，但可以认为，此时的容量还是式（4-7）所示的平均容量，只不过发送功率沿时间进行了最佳的分配。我们将平均功率受限下的衰落信道容量定义为

$$C=\max_{P(\gamma):\int P(\gamma)p(\gamma)\mathrm{d}\gamma=\bar{P}}\int_0^\infty B\log_2\left(1+\frac{P(\gamma)\gamma}{\bar{P}}\right)p(\gamma)\,\mathrm{d}\gamma \tag{4-9}$$

文献[20]证明，由式（4-9）所确定的容量是可以实现的，而且，若传输速率大于该容量，则误码率必不能任意小。文献[20]中的证明过程用到了如图4-4所示的多路输入多路输出的时间分集系统。具体来说，首先将衰落值γ量化为有限集合$\{\gamma_j:1\leqslant\gamma_j\leqslant N\}$。针对每个$\gamma_j$设计一组AWGN信道下的编译码器。若$i$时刻的信噪比$\gamma[i]\approx\gamma_j$，信道的输入输出就连接到第$j$路编译码器上。每一路对应的码字以多路复用的方式送入信道并在接收端解复用输出。第j路上的平均发送功率是$P(\gamma_j)$，数据传输速率$R_j=C_j$是信噪比为$P(\gamma_j)\gamma_j/\bar{P}$的时不变AWGN信道的容量。这种方法将时变信道分解为一组并行的时不变信道，第j个信道只在$\gamma[i]\approx\gamma_j$时才被使用。各路的传输速率按其出现概率$p(\gamma_j)$做加权和就得108到信道的平均传输速率。这样就产生了式（4-9）的平均信道容量。

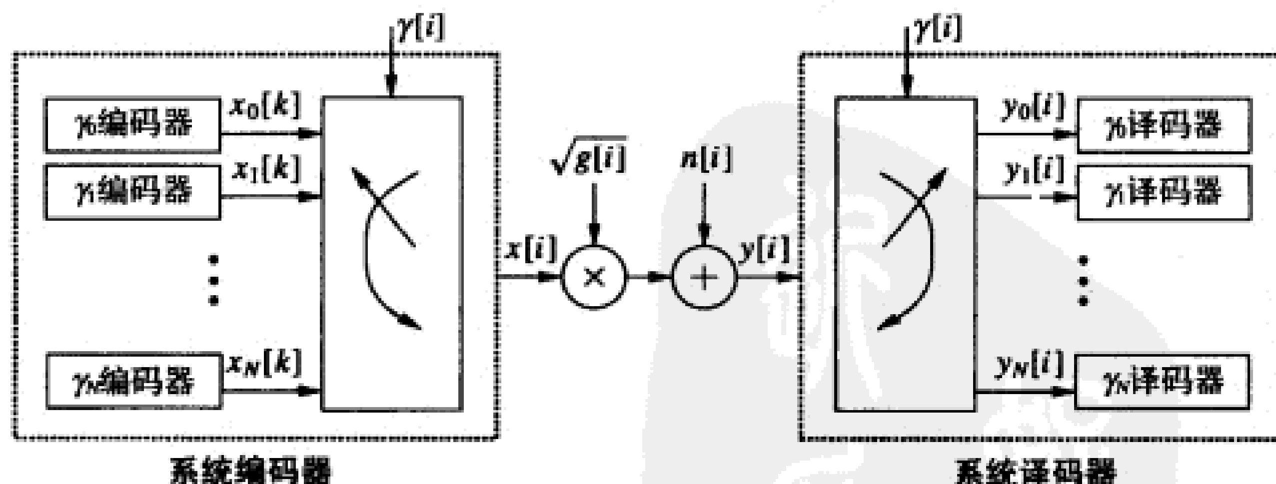


图4-4 编译码多路复用

为了找到最优的功率分配方式$P(\gamma)$，构造拉格朗日条件式

$$J(P(\gamma))=\int_0^\infty B\log_2\left(1+\frac{\gamma P(\gamma)}{\bar{P}}\right)p(\gamma)\,\mathrm{d}\gamma-\lambda\int_0^\infty P(\gamma)p(\gamma)\,\mathrm{d}\gamma \tag{4-10}$$

1. Wolfowitz的结果中，γ取值于有限集合，但可推广到无限集合[20]。

求导并令导数为零：

$$\frac{\partial J(P(\gamma))}{\partial P(\gamma)}=\left[\left(\frac{B/\ln(2)}{1+\gamma P(\gamma)/\bar{P}}\right)\frac{\gamma}{\bar{P}}-\lambda\right]p(\gamma)=0 \tag{4-11}$$

按约束条件$P(\gamma)>0$求解$P(\gamma)$即可得到使式（4-9）最大的最佳功率分配为

$$\frac{P(\gamma)}{\bar{P}}=\begin{cases}1/\gamma_0-1/\gamma & \gamma\geqslant\gamma_0\\ 0 & \gamma<\gamma_0\end{cases} \tag{4-12}$$

其中γ_0为中断门限。当$\gamma[i]$低于γ_0时不进行数据传输，仅当$\gamma[i]$大于γ_0时才使用信道。将式（4-12）代入式（4-9）得到信道容量公式为

$$C=\int_{\gamma_0}^{\infty}B\log_2\left(\frac{\gamma}{\gamma_0}\right)p(\gamma)\,\mathrm{d}\gamma \tag{4-13}$$

上述论证过程中的多路复用特性表明，式（4-13）的容量可通过时变的传输速率来达到，瞬时信噪比γ对应的速率为$B\log_2(\gamma/\gamma_0)$。由于γ_0是固定的，所以瞬时信噪比增加时，对应的传输速率也要增加。注意这种复用方案并不是实现式（4-13）的唯一途径，也可以通过实时调整发送功率但固定传输速率来实现[21]。在4.2.6节中我们将看到，对于瑞利衰落信道，这个容量有可能超过相同平均发送功率的AWGN信道，而在CSI仅为接收端所知的情况，容量总会因为衰落而减小。

注意对于式（4-12）所示的最佳功率分配与衰落分布$p(\gamma)$的关系仅体现在中断门限γ_0上。γ_0可根据功率限制条件求得。由于功率取最大时的容量总是最大的，所以可对式（4-8）取等号，再重新整理为

$$\int_0^{\infty}\frac{P(\gamma)}{\bar{P}}p(\gamma)\,\mathrm{d}\gamma=1 \tag{4-14}$$

将式（4-12）代入，则γ_0必然满足

$$\int_{\gamma_0}^{\infty}\left(\frac{1}{\gamma_0}-\frac{1}{\gamma}\right)p(\gamma)\,\mathrm{d}\gamma=1 \tag{4-15}$$

此式只与分布$p(\gamma)$有关。对于一般的连续概率密度函数$p(\gamma)$，不存在γ_0的闭式解，只能通过数值计算求得[22]。 109

由于γ是时变的，所以式（4-12）所示的最优功率分配方法是时域的注水法功控，如图4-5所示。图中的曲线示出了不同瞬时信噪比γ所分配的功率。注水法的原理是，曲线$1/\gamma$表示一个碗的底部，将功率不断注入这个碗，直到达到$1/\gamma_0$处。那么，对于给定的γ所分配的功率（$1/\gamma_0-1/\gamma$）就是此处的水深。注水法的直观解释就是尽量充分利用好的信道条件：信道好（γ大）的时候，多分配功率，用高速率传输；如果信道质量变差（γ小），就减少分配功率，并用低速率传输。如果瞬时信噪比低于中断门限γ_0，则不进行传输。文献[23~24]提出了根据这一原则进行的自适应调制编码技术，本书第9章将对此进行讨论。

注意前面论证式（4-9）中所用的多路复用方法也可以用于任意的功率分配方案。就是说，对于任何受限于平均功率$\bar{P}$的功控方案$P(\gamma)$，一定能够以任意小的误码率实现

容量

$$C=\int_0^{\infty} B\log_2\left(1+\frac{P(\gamma)\gamma}{\bar{P}}\right)p(\gamma)\,\mathrm{d}\gamma \tag{4-16}$$

当然这个容量不可能超过式（4-9）中的容量，因为式（4-9）中的功率分配是能使容量最大化的最优功率分配。不过在某些情况下，我们可能会出于一些比容量更重要的因素而采用次佳的功控方案。在接下来的两节里，我们将讨论两种次佳功控方案，它们的传输速率是恒定的，而式（4-9）对应的是变速率传输。

110

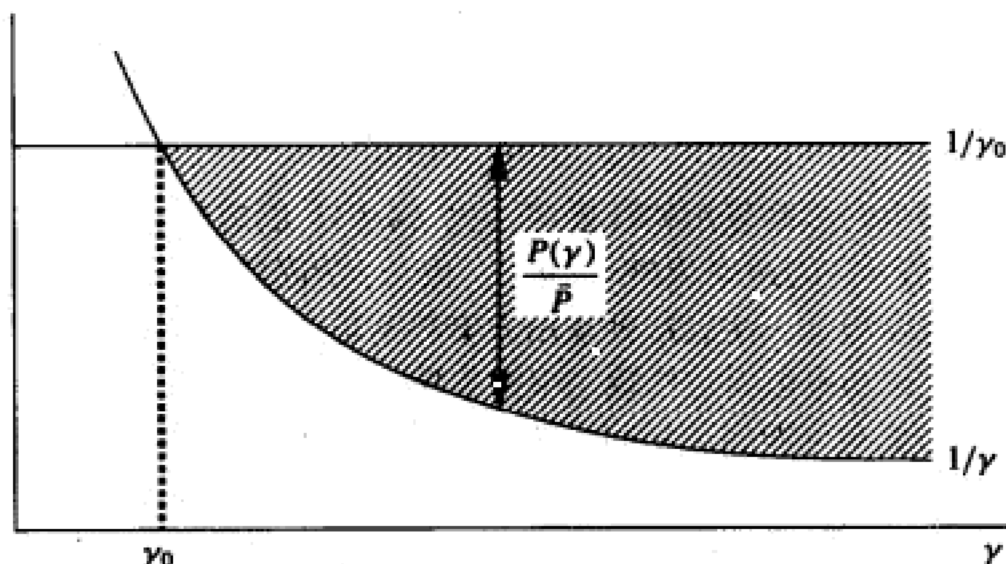


图4-5　最佳功率分配：注水法

例4.4：考虑与前例相同的信道，带宽为30kHz，有三个可能的接收信噪比：出现$\gamma_1=0.8333$的概率是$p_1=0.1$，出现$\gamma_2=83.33$的概率是$p_2=0.5$，出现$\gamma_3=333.33$的概率是$p_3=0.4$。假设发送端和接收端都知道瞬时的CSI，求此信道的遍历容量。

解：我们知道最佳功控方案为注水法，先来确定中断门限γ_0。式（4-15）对应到离散随机变量是

$$\sum_{\gamma_i>\gamma_0}\left(\frac{1}{\gamma_0}-\frac{1}{\gamma_i}\right)p(\gamma_i)=1 \tag{4-17}$$

先假设所有信道状态都会用来传输，即假设$\gamma_0\leqslant\min_i\gamma_i$，然后验证求得的$\gamma_0$是否小于最差信道状态的信噪比。如果不是将产生矛盾，需要重新假设至少有一个信道状态是不传输数据的，再重新进行计算。将本题条件代入式（4-17）得

$$\sum_{i=1}^{3}\frac{p(\gamma_i)}{\gamma_0}-\sum_{i=1}^{3}\frac{p(\gamma_i)}{\gamma_i}=1$$

$$\Longrightarrow\ \frac{1}{\gamma_0}=1+\sum_{i=1}^{3}\frac{p(\gamma_i)}{\gamma_i}=1+\left(\frac{0.1}{0.8333}+\frac{0.5}{83.33}+\frac{0.4}{333.33}\right)=1.13$$

解得$\gamma_0=1/1.13=0.89>0.8333=\gamma_1$。因为只有当信噪比大于$\gamma_0$时才进行传输，但前面假设所有信噪比下都传输，出现矛盾。因此，我们假设信噪比为γ_1时不传输，重新计算γ_0。此时式（4-17）成为

$$\sum_{i=2}^{3}\frac{p(\gamma_i)}{\gamma_0}-\sum_{i=2}^{3}\frac{p(\gamma_i)}{\gamma_i}=1$$

$$\Longrightarrow \frac{0.9}{\gamma_0}=1+\sum_{i=2}^{3}\frac{p(\gamma_i)}{\gamma_i}=1+\left(\frac{0.5}{83.33}+\frac{0.4}{333.33}\right)=1.0072$$

解得$\gamma_0=0.89$，满足$\gamma_1<\gamma_0<\gamma_2$。信道容量为

$$C=\sum_{i=2}^{3}B\log_2\left(\frac{\gamma_i}{\gamma_0}\right)p(\gamma_i)$$

$$=30\,000\left(0.5\log_2\frac{83.33}{0.89}+0.4\log_2\frac{333.33}{0.89}\right)=200.82\text{ kbit/s}$$

与例4.3相比，本例的结果比接收端已知CSI情况下的容量只是少许提高，同样比相同平均信噪比的AWGN信道容量低很多。这是因为本例的平均信噪比相对较高。在低信噪比的平衰落信道中，通过利用衰落信道中偶尔出现的好的信道状态，我们可以获得比相同平均信噪比下的AWGN信道更高的容量。

2. 零中断容量与信道反转

下面我们考虑一种次佳的功率分配方式，它依靠发送端已知的CSI使接收端保持恒定接收功率，即将信道衰落进行反转。这样对编译码器来说，信道呈现为时不变的AWGN信道。这种功率分配叫做信道反转（channel inversion），由$P(\gamma)/\overline{P}=\sigma/\gamma$确定，其中$\sigma$是满足式（4-8）发送功率约束条件下，可以维持的恒定接收信噪比，σ满足$\int(\sigma/\gamma)p(\gamma)\mathrm{d}\gamma=1$，即$\sigma=1/E[1/\gamma]$。

111

信道反转后衰落信道的容量就是信噪比为σ的AWGN信道的容量

$$C=B\log_2[1+\sigma]=B\log_2\left[1+\frac{1}{E[1/\gamma]}\right] \tag{4-18}$$

达到该容量的传输方式是用针对信噪比为σ的AWGN设计的固定速率的编译码器实现的。其优点是可以不用考虑信道的状态而以固定速率进行传输。由于在所有信道状态下传输速率不变，传输过程不会发生中断，所以式（4-18）的容量又叫做**零中断容量**（zero-outage capacity）。由于实际中的编码技术在AWGN信道中已经能获得接近容量的数据速率，所以零中断容量是可以近似实现的。

在衰落严重的环境中，零中断容量相对于香农容量有大幅度的速率下降。例如在瑞利衰落下$E[1/\gamma]$是无穷大，因此式（4-18）给出的容量为零。扩频系统经常使用信道反转以克服远近干扰不平衡[25]。信道反转系统实现简单，编译码器可以不考虑衰落而只按AWGN信道设计。

例4.5：考虑与前例相同的信道，带宽为30kHz，有三个可能的接收信噪比：出现$\gamma_1=0.8333$的概率是$p_1=0.1$、出现$\gamma_2=83.33$的概率是$p_2=0.5$、出现$\gamma_3=333.33$的概率是$p_3=0.4$。假设收发都已知CSI，求此信道的零中断容量。

解：零中断容量是$C=B\log_2[1+\sigma]$，其中$\sigma=1/E[1/\gamma]$。由于

$$E\left[\frac{1}{\gamma}\right]=\frac{0.1}{0.8333}+\frac{0.5}{83.33}+\frac{0.4}{333.33}=0.1272$$

我们得到$C=30\,000\log_2(1+1/0.1272)=94.43$kbit/s。该容量还不到注水法功控得到的香农容量的一半。

3. 中断容量与截断式信道反转

衰落信道的零中断容量显著低于香农容量的原因在于它要在所有衰落状态下维持恒定的传输速率。如果在特别差的信道状态下停止发送数据（称此时为中断状态），就可以提高非中断状态下的传输速率，从而显著提高容量。**中断容量**（outage capacity）定义为非中断的概率和非中断状态下可维持的最大传输速率的乘积。可通过*截断式信道反转*的功控方案来实现中断容量，它只在信道条件好于中断门限γ_0时才补偿衰落：

$$\frac{P(\gamma)}{\bar{P}}=\begin{cases}\sigma/\gamma & \gamma\geqslant\gamma_0\\ 0 & \gamma<\gamma_0\end{cases} \tag{4-19}$$

其中γ_0由中断率$P_{out}=P(\gamma<\gamma_0)$决定。因为只在$\gamma\geqslant\gamma_0$时才利用信道，从式（4-8）的功率限制条件可得$\sigma=1/E_{\gamma_0}[1/\gamma]$，其中

112

$$E_{\gamma_0}\left[\frac{1}{\gamma}\right]\triangleq\int_{\gamma_0}^{\infty}\frac{1}{\gamma}p(\gamma)\,\mathrm{d}\gamma \tag{4-20}$$

给定中断概率P_{out}及相应的γ_0时，中断容量由下式给出

$$C(P_{out})=B\log_2\left(1+\frac{1}{E_{\gamma_0}[1/\gamma]}\right)p(\gamma\geqslant\gamma_0) \tag{4-21}$$

我们还可以通过优化γ_0来得到**最大中断容量**（maximum outage capacity）：

$$C=\max_{\gamma_0}B\log_2\left(1+\frac{1}{E_{\gamma_0}[1/\gamma]}\right)p(\gamma\geqslant\gamma_0) \tag{4-22}$$

最大中断容量仍然会比式（4-13）给出的香农容量小，因为截断式信道反转仍然是一种次佳的传输方案。不过，采用信道反转或截断式信道反转的系统实现起来要比采用注水法的系统简单。

例4.6：考虑与前例相同的信道，带宽为30kHz，有三个可能的接收信噪比：出现$\gamma_1=0.8333$的概率是$p_1=0.1$、出现$\gamma_2=83.33$的概率是$p_2=0.5$、出现$\gamma_3=333.33$的概率是$p_3=0.4$。求$\gamma_0=0.84$及$\gamma_0=83.4$下的中断概率及中断容量并比较这两个值条件下的中断容量大小。

解：当$\gamma_0=0.84$时，只有当信噪比为γ_2或γ_3时才进行传输，所以$E_{\gamma_0}[1/\gamma]=\sum_{i=2}^{3}p_i/\gamma=0.5/83.33+0.4/333.33=0.0072$。中断容量为$C=B\log_2(1+1/E_{\gamma_0}[1/\gamma])p(\gamma\geqslant\gamma_0)=192.457$kbit/s。当$\gamma_0=83.34$时，只有信噪比为$\gamma_3$时才进行传输，$E_{\gamma_0}[1/\gamma]=p(\gamma_3)/\gamma_3=0.0012$。中断容量为$C=B\log_2(1+1/E_{\gamma_0}[1/\gamma])p(\gamma\geqslant\gamma_0)=116.45$kbit/s。可见，在$\gamma_0=0.84$时中断容量比较大。虽然$\gamma_3$比$\gamma_2$大很多，但它的出现概率只有40%，所以只在最佳状态下才使用信道的效率并不高。

4.2.5 接收分集的信道容量

接收分集是无线通信中一项著名的抗衰落技术。它可以降低衰落引起的信道起伏，

使信道接近于AWGN。本书第7章将详细讨论接收分集及其性能。由于接收分集可减小衰落的影响，我们很自然会想到它是否会增加衰落信道的容量。信道容量的计算首先需要知道给定分集合并方式下接收信噪比的分布$p(\gamma)$。一旦知道了分布，就可以代入前面给出的任何一个容量公式进行计算，得到相应的分集合并下的信道容量。用哪个计算公式取决于对CSI的假设，例如收发端都已知CSI时应使用式（4-13）进行计算。对于不同的CSI假设，文献[22]计算了最大比合并及选择合并下的信道容量，其结果表明，收发两端都已知CSI时的容量大于只有接收端已知CSI的情况，更大于信道反转功控的情况。随着分集天线数的增加，前述的这些不同定义的信道容量数值逐步接近。这是因为天线数量越多，信道越接近AWGN，而在AWGN信道中这些容量公式的计算结果基本相同。 113

近年来，收发端都采用多个天线的系统成为一个研究热点。这主要是因为文献[26~28]的突破性成果指出：多入多出衰落信道（MIMO信道）的信道容量是单天线信道的M倍，其中$M=\min(M_t, M_r)$，M_t、M_r分别是发送和接收天线的数目。我们将在第10章讨论多天线系统的容量。

4.2.6　容量对比

本节中我们将通过具体的信道实例对比前面介绍过的几种信道容量。图4-6、图4-7、图4-8分别给出了对数正态衰落（标准差$\sigma=8$dB）、瑞利衰落、Nakagami衰落（Nakagami参数$m=2$）下，由式（4-4）、式（4-13）、式（4-18）、式（4-22）确定的容量随平均接收信噪比变化的曲线。$m=2$的Nakagami分布近似等价于两重接收分集的瑞利衰落。为便于比较，图中还给出了相同平均信噪比下AWGN信道的容量。注意在对数正态衰落的图中，横坐标是分贝信噪比的平均值μ_{dB}，不是平均信噪比的分贝值$10\log_{10}\mu$，二者的关系如第2章的式（2-45）所示，是$10\log_{10}\mu=\mu_{dB}+\sigma_{dB}^2\ln(10)/20$。

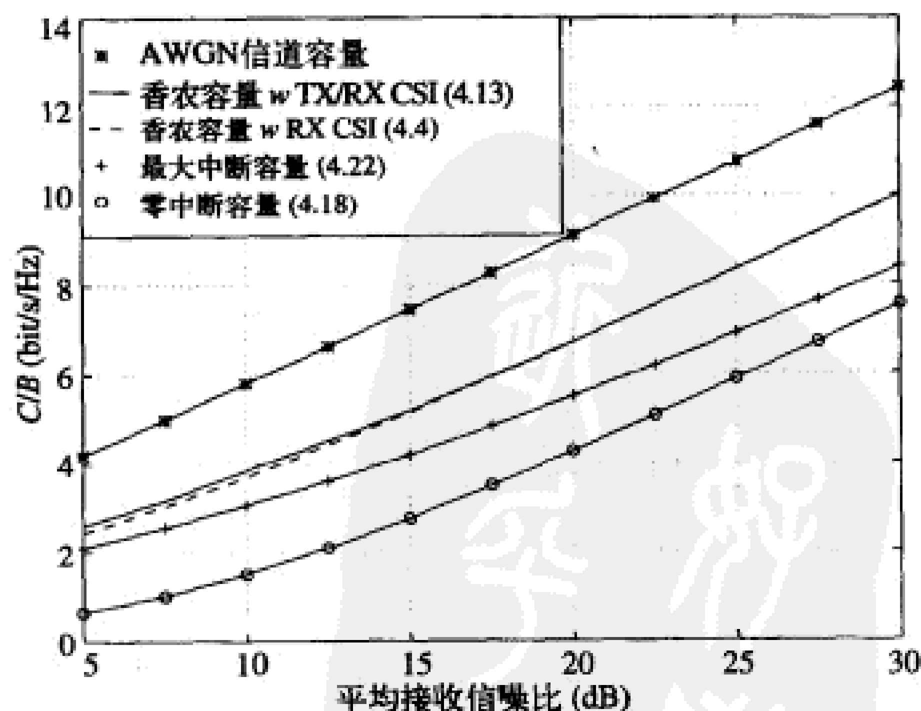


图4-6　对数正态衰落下的信道容量

比较这些结果，有几点值得注意。首先，从图中来看，AWGN信道的容量比任何

情况下衰落信道容量都大。不过在信噪比较低时，AWGN信道的容量与收发两端都已知CSI的容量非常接近。实际上，对于更低的信噪比（小于0dB），收发两端都已知CSI的信道容量还要更大一些。这是因为AWGN信道中的信噪比始终是一个低值，它限制了信道容量。但对于同样低的平均信噪比的衰落信道，衰落分布的无限性会导致瞬时的高信噪比。因此，若将功率主要集中在这些偶尔出现的高信噪比上，并用高速率传输，就能产生出比相同低平均信噪比的AWGN信道更高的容量。

114

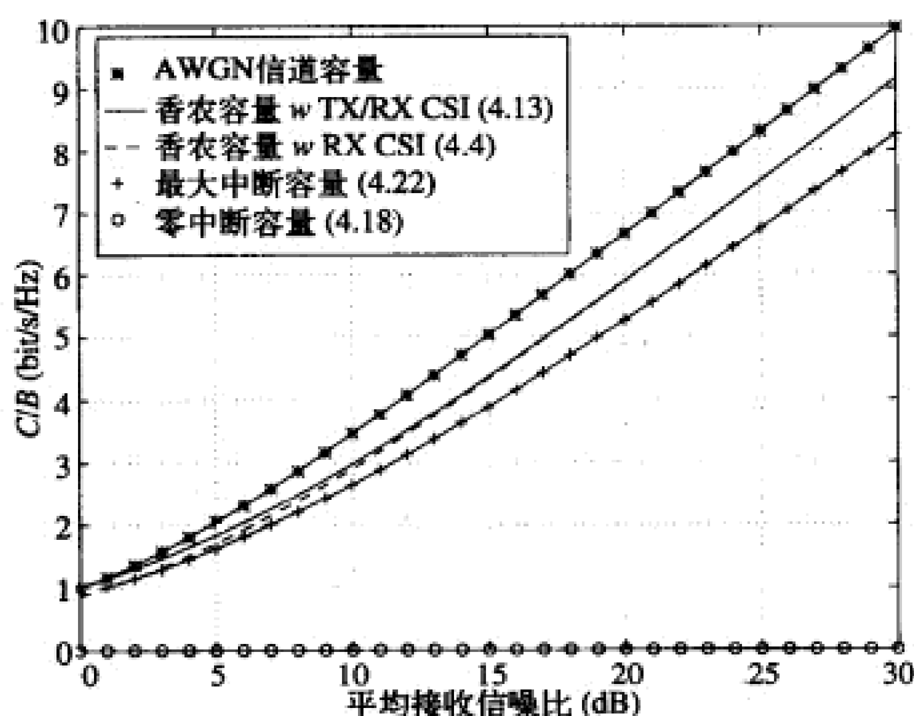


图4-7 瑞利衰落的容量

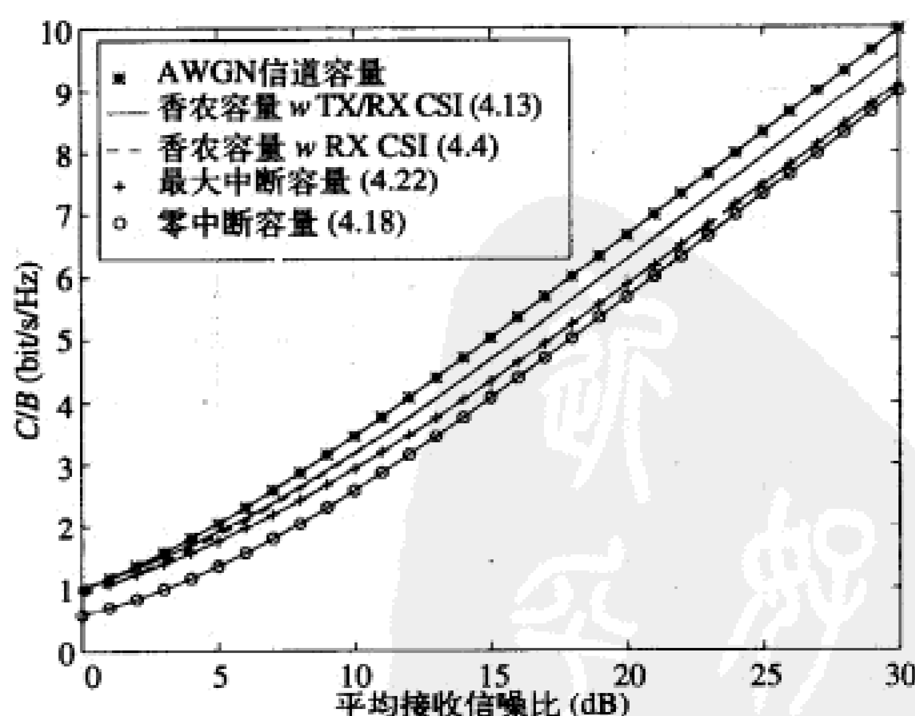


图4-8 Nakagami衰落（$m=2$）的容量

Nakagami衰落的参数m表征衰落的严重程度，$m=1$对应瑞利衰落，$m=\infty$对应无衰落的AWGN信道。因此对照图4-7与图4-8可见，当衰落程度减小时（从瑞利衰落到$m=$

2的Nakagami衰落)，图中的各种容量曲线之间的差别在减小，相应的容量值也逐渐接近于AWGN信道。

115

在各种情况下，收发端都已知CSI时的信道容量式（4-13）与只有接收端已知CSI的信道容量式（4-4）差别很小。收发两端已知CSI但无功控的容量式（4-7）与接收端已知CSI的容量式（4-4）是相同的，这说明，只要自适应根据信道调整传输速率，那么进一步自适应调整功率对容量的提高几乎可以忽略不计。也说明，相对于接收端已知CSI的信道容量来说，发送端获得CSI基本不增加容量。在严重的衰落情况下（如瑞利衰落及对数正态衰落），最大中断容量相比于香农容量有1~5dB的损失，而零中断容量相比香农容量损失巨大。但在衰落不太严重时（$m=2$的Nakagami衰落），中断容量、零中断容量与香农容量相差3dB以内，与AWGN信道相差4dB以内。这个差别将随着衰落趋向于消失（Nakagami衰落的$m\to\infty$）而进一步减小。

我们可以将这些结果看作是容量和复杂性之间的折中。收发两端都已知CSI的自适应方案需要增加发送端的复杂度，而且一般还需要有一个从接收端到发送端的反馈信道，但接收端译码部分要相对简单一些。非自适应的方案发送端比较简单，但编码设计依赖于信道的相关性（往往是未知的），且译码的复杂程度正比于信道的去相关时间。信道反转与截断式信道反转方案可采用AWGN信道的编码，是最容易实现的，但在衰落较严重时会有很大的容量损失。

一般来说，香农容量分析并不能告诉我们，该怎样为实际系统设计自适应或非自适应的技术。文献[24]研究了自适应网格编码MQAM调制可达到的速率，采用简单的四状态格码结合自适应六星座MQAM调制，可以做到离图4-6、图4-7所示的容量式（4-9）只差7dB。更复杂的编码结合自适应发送可以进一步缩小这个差距。

4.3 频率选择性衰落信道的容量

本节我们研究频率选择性衰落信道的香农容量。我们首先考虑时不变频率选择性衰落信道，其容量分析与平衰落信道类似，只是把时间轴换成了频率轴。然后我们讨论时变频率选择性衰落信道。

4.3.1 时不变信道

考虑图4-9所示的时不变信道，其频率响应为$H(f)$。假设总发送功率上限为P。对于时不变信道，一般可以假设$H(f)$对接收端和发送端都是已知的。其他信道信息假设下时不变信道的容量分析见文献[19, 21]。

116

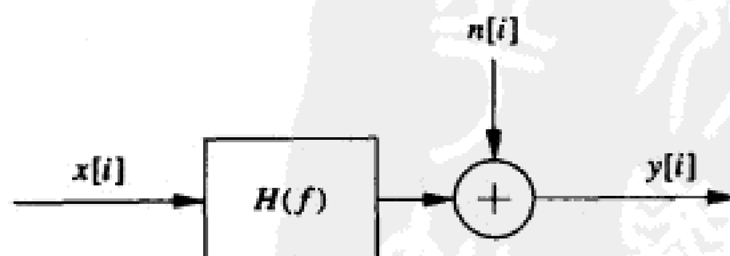


图4-9 时不变频率选择性衰落信道

PDG

首先假设$H(f)$是分块衰落（block fading）的，因此整个频带可以分割为带宽为B的许多子信道，如图4-10所示。在每个子信道中$H(f)=H_j$是恒定的。这样，频率选择性信道变成了一组并行的AWGN信道，其中第j个信道的信噪比为$|H_j|^2P_j/(N_0B)$，P_j是在功率限制$\sum_j P_j \leqslant P$的约束下分配给该子信道的功率。

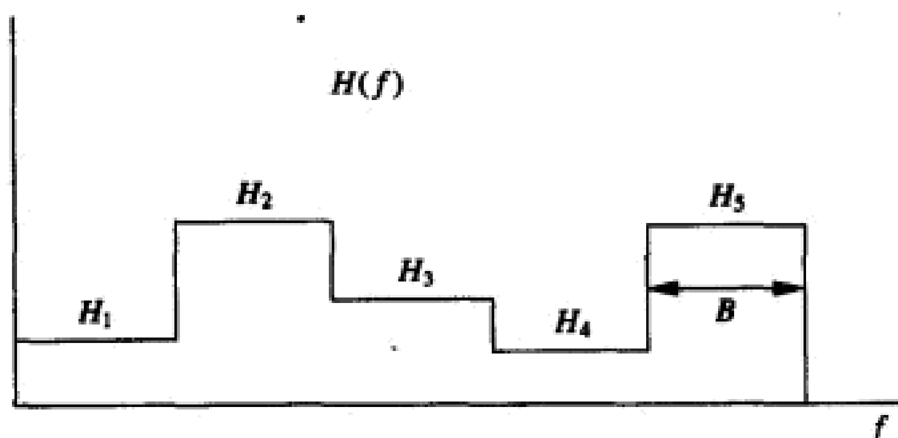


图4-10 分块频率选择性衰落

信道容量是经过了最佳功率分配后，所有子信道的速率之和[5-6]：

$$C = \sum_{\max P_j: \sum_j P_j \leqslant P} B \log_2\left(1 + \frac{|H_j|^2 P_j}{N_0 B}\right) \tag{4-23}$$

注意这里的容量及最佳功率分配和平衰落信道下的情形相似，只不过这里的功率和速率分配是沿频域以确定的方式进行的，而在平衰落信道中功率和速率的分配是在时域以概率的方式进行的。同样可以通过拉格朗日法得到最佳的功率分配，即按如下的注水法功控分配：

$$\frac{P_j}{P} = \begin{cases} 1/\gamma_0 - 1/\gamma_j & \gamma_j \geqslant \gamma_0 \\ 0 & \gamma_j < \gamma_0 \end{cases} \tag{4-24}$$

其中γ_0为中断门限，$\gamma_j=|H_j|^2P/(N_0B)$是假设将所有功率都分给这个子信道时，第j个子信道的信噪比。图4-11示出了这种最佳功率分配。将功控公式（4-24）代入功率受限公式可得到中断门限γ_0必须满足

117

$$\sum_j \left(\frac{1}{\gamma_0} - \frac{1}{\gamma_j}\right) = 1 \tag{4-25}$$

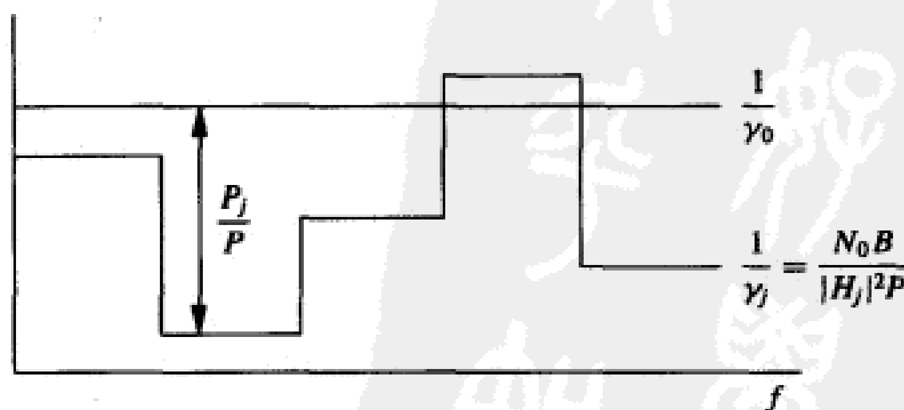


图4-11 分块频率选择信道的注水法功控

信道容量为

$$C = \sum_{j:\gamma_j \geq \gamma_0} B \log_2\left(\frac{\gamma_j}{\gamma_0}\right) \tag{4-26}$$

在不同子信道上分配不同功率并以不同速率传输来获得这个信道容量。多载波调制中的自适应负载也使用了类似的技术，本书12.3节将对此做更多的讨论。

当$H(f)$连续时，功率受限于P的信道容量问题与分块衰落时的情形类似，通过一些复杂的数学推导，可以证明信道容量为

$$C = \max_{P(f): \int P(f)\,\mathrm{d}f \leq P} \int \log_2\left(1 + \frac{|H(f)|^2 P(f)}{N_0}\right) \mathrm{d}f \tag{4-27}$$

积分号里面的部分可以看作是在频率f、带宽$\mathrm{d}f$，功率分配$P(f)$，信道增益$|H(f)|^2$条件下的信道容量。这一结果的正式证明利用了Karhunen-Loeve展开，把信道的$h(t)$等价为一组独立并行信道（见文献[5]的8.5节）。另一种证明方法[29]是用离散傅里叶变换（discrete Fourier transform）将信道进行并行分解，本书12.4节在讨论多载波调制的离散实现时，也将提到类似的做法。

可通过拉格朗日法求得使式（4-27）最大的频域最佳功率分配，结果是频域的注水法功控：

$$\frac{P(f)}{P} = \begin{cases} 1/\gamma_0 - 1/\gamma(f) & \gamma(f) \geq \gamma_0 \\ 0 & \gamma(f) < \gamma_0 \end{cases} \tag{4-28}$$

其中$\gamma(f) = |H(f)|^2 P/(N_0 B)$。由此得信道容量为

$$C = \int_{f:\gamma(f) \geq \gamma_0} \log_2\left(\frac{\gamma(f)}{\gamma_0}\right) \mathrm{d}f \tag{4-29}$$

118

例4.7：考虑一个时不变频率选择性分块衰落信道，它由三个带宽为1MHz的子信道组成，频率响应分别为$H_1 = 1$、$H_2 = 2$、$H_3 = 3$。发送端功率上限为10mW，噪声的功率谱密度为$N_0/2$，$N_0 = 10^{-9}$W/Hz。求信道的香农容量及对应的最佳功率分配。

解：先求出每个子信道的信噪比$\gamma_j = |H_j|^2 P/(N_0 B)$，得到$\gamma_1 = 10$、$\gamma_2 = 40$、$\gamma_3 = 90$。中断门限$\gamma_0$要满足式（4-25）。假设在所有子信道上都分配功率，则

$$\frac{3}{\gamma_0} = 1 + \sum_j \frac{1}{\gamma_j} = 1.14 \implies \gamma_0 = 2.64 < \gamma_j \ \forall j$$

由于γ_0小于所有子信道的信噪比，所以假设成立，因此中断门限就是$\gamma_0 = 2.64$。对应的信道容量为$C = \sum_{j=1}^{3} B\log_2(\gamma_j/\gamma_0) = 10.93$Mbit/s。

4.3.2 时变信道

时变频率选择性衰落信道的模型与图4-9类似，只不过要把$H(f)$换写为$H(f, i)$，表示信道既随频率变化，也随时间变化。由于码间干扰（intersymbol interference, ISI）的影响，即使收发端完全已知信道的瞬时$H(f, i)$，容量分析也是非常困难的。收发端都已

知CSI时，最佳自适应方案需要考虑：(a) 信道对已发送符号序列的影响，以及 (b) 这些符号序列的ISI对后续传输的影响[30]。就一般情形来说，时变频率选择性信道的容量是未知的，但存在上下界和极限公式[30-31]。

119

作为一种近似，可将带宽为B的信道按相干带宽B_c分割为多个子信道，如图4-12所示。然后假设各子信道是相互独立的时变平衰落信道，对于第j个子信道有$H(f,i)=H_j[i]$。

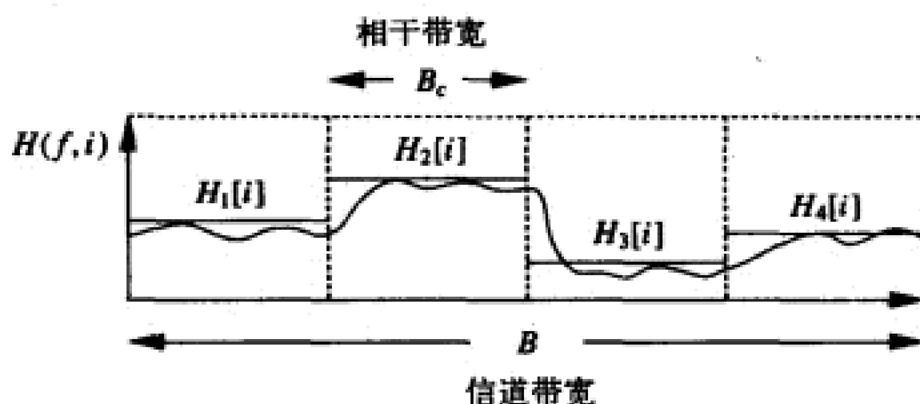


图4-12 频率选择性衰落下的信道划分

在这些假设下，我们根据总上限功率 $\bar{P}$ 给每个子信道分配平均功率 $\bar{P}_j$，然后得到平均功率为 $\bar{P}_j$的各个平衰落子信道的容量。由于各子信道相互独立，我们可以得到沿时间和频率平均的总功率约束条件下的总信道容量，它是各窄带平衰落子信道容量之和：

$$C=\max_{\{\bar{P}_j\}:\sum_j\bar{P}_j\leqslant\bar{P}}\sum_j C_j(\bar{P}_j) \tag{4-30}$$

其中 $C_j(\bar{P}_j)$是平均功率为 $\bar{P}_j$，带宽为B_c的平衰落子信道的容量，按照不同的CSI假设及不同的功率分配策略，它可由式 (4-13)、式 (4-4)、式 (4-18) 或式 (4-22) 算出。如果仅接收端已知CSI，我们也可以定义 $C_j(\bar{S}_j)$为带中断的容量。

考虑到收发端都已知CSI时的香农容量是其他所有CSI假设及功控策略下信道容量的上界，所以接下来我们将只考虑这种情形。我们知道，如果确定了每个子信道的平均功率，则子信道上沿时间的最佳功控满足注水法功控。现在我们设想把总平均功率分给各个子信道时，也服从注水法功控，即好的子信道分得更大的平均功率。这样就构成了时频二维的注水法功控。下面我们来介绍这种二维注水法功控以及相应的香农容量。

定义$\gamma_j[i]=|H_j[i]|^2\bar{P}/(N_0B_c)$为将全部功率$\bar{P}$都分配给时刻$i$第$j$个子信道上时所得到的瞬时信噪比。令$P_j(\gamma_j)$可以随$\gamma_j[i]$变化，则收发端都已知CSI时，对时域（体现在$\gamma_j[i]=\gamma_j$上）和频域（体现为子信道号$j$）进行最佳功控得到的香农容量为：

$$C=\max_{P_j(\gamma_j):\sum_j\int_0^\infty P_j(\gamma_j)p(\gamma_j)\mathrm{d}\gamma_j\leqslant\bar{P}}\sum_j\int_0^\infty B_c\log_2\left(1+\frac{P_j(\gamma_j)\gamma_j}{\bar{P}}\right)p(\gamma_j)\mathrm{d}\gamma_j \tag{4-31}$$

为了得到最佳功控方案$P_j(\gamma_j)$，可建立如下拉格朗日条件式

$$J(P_j(\gamma_j))=\sum_j\int_0^\infty B_c\log_2\left(1+\frac{P_j(\gamma_j)\gamma_j}{\bar{P}}\right)p(\gamma_j)\mathrm{d}\gamma_j-\lambda\sum_j\int_0^\infty P_j(\gamma_j)p(\gamma_j)\mathrm{d}\gamma_j \tag{4-32}$$

注意上式与平衰落信道下的拉格朗日条件式（4-10）类似，只是通过对子信道求和增加了频率维数。对上式求导并令其为零，则除了相关的子信道及信噪比之外，其余项都被消除：

$$\frac{\partial J(P_j(\gamma_j))}{\partial P_j(\gamma_j)} = \left[\left(\frac{B_c/\ln(2)}{1+\gamma_j P_j(\gamma_j)/\bar{P}}\right)\frac{\gamma_j}{\bar{P}} - \lambda\right]p(\gamma_j) = 0 \tag{4-33}$$

120

由此解得与平衰落情形相同的注水法功控：

$$\frac{P_j(\gamma_j)}{\bar{P}} = \begin{cases} 1/\gamma_0 - 1/\gamma_j & \gamma_j \geqslant \gamma_0 \\ 0 & \gamma_j < \gamma_0 \end{cases} \tag{4-34}$$

其中的中断门限γ_0可通过时域和频域上的总功率约束求得：

$$\sum_j \int_0^{\infty} P_j(\gamma_j)p(\gamma_j)\,\mathrm{d}\gamma_j = \bar{P} \tag{4-35}$$

因此式（4-34）所示的功率分配是有同一γ_0的二维注水法功控。将式（4-35）除以$\bar{P}$并代入式（4-34）得γ_0必须满足

$$\sum_j \int_{\gamma_0}^{\infty}\left(\frac{1}{\gamma_0} - \frac{1}{\gamma_j}\right)p(\gamma_j)\,\mathrm{d}\gamma_j = 1 \tag{4-36}$$

有趣的是，二维注水法功控的中断门限γ_0对所有子信道是相同的。这说明，即使各子信道上的衰落分布或者平均功率不同，只要瞬时信噪比低于γ_0，这些子信道都会停止传输。将最佳功控（4-35）代入信道容量公式（4-31）得

$$C = \sum_j \int_{\gamma_0}^{\infty} B_c \log_2\left(\frac{\gamma_j}{\gamma_0}\right)p(\gamma_j)\,\mathrm{d}\gamma_j \tag{4-37}$$

习题

4.1 AWGN信道的容量是$C=B\log_2(1+P/(N_0B))$，其中P为接收信号功率，B为信号带宽，$N_0/2$为噪声功率谱密度。求带宽$B\to\infty$时，信道容量关于P的函数。

4.2 某AWGN信道的带宽是50MHz，接收信号功率是10mW，噪声的功率谱密度是$N_0/2$，其中$N_0=2\times10^{-9}$W/Hz。求接收功率和带宽分别增加一倍时信道容量的增量。

4.3 假设在AWGN信道中，有两个用户同时向一个接收端发送信号。这种情况如同蜂窝系统中多个用户同时向基站发信号。假设接收机收到的每个用户的功率都为10mW，接收机的噪声功率是0.1mW，每个的用户带宽都是20MHz。

(a) 假设接收机先对第一个用户进行解码。这时第二个用户的信号就被当作噪声（假设第二个用户的信号与AWGN信号具有相同统计特性）。那么在这种有干扰和噪声的情况下，第一个用户的信道容量是多少？

121

(b) 假设对第一个用户解码后，接收机重新对其进行编码并从接收信号中减去第一个用户信号。这样在对第二个用户解码时就没有了第一个用户的干扰。求此时对第二个用户而言的信道容量。

注：在第14章中我们将会看到，对于多个独立的发射机同时向一个接收机发送信号的情

况，这种从接收信号中顺序减去已知信号的方法可以达到香农容量。

4.4 设某平衰落信道的带宽为20MHz。若将发射功率固定为$\overline{P}$，则接收信噪比有以下六个可能值：$\gamma_1=20$dB、$\gamma_2=15$dB、$\gamma_3=10$dB、$\gamma_4=5$dB、$\gamma_5=0$dB、$\gamma_6=-5$dB，对应的出现概率为$p_1=p_6=0.1$、$p_2=p_4=0.15$、$p_3=p_5=0.25$。假设只有接收端已知CSI。

(a) 求此信道的香农容量。

(b) 画出带中断的容量随中断率变化的曲线（$0\leqslant P_{out}<1$），并求出能被正确接收的最大平均速率（最大C_{out}）。

4.5 设有一平衰落信道，发送功率为$\overline{P}$，接收信噪比有以下四种可能取值：$\gamma_1=30$dB、$\gamma_2=20$dB、$\gamma_3=10$dB及$\gamma_4=0$dB，相应的出现概率为$p_1=p_4=0.2$、$p_2=p_3=0.3$。假设收发端都已知CSI。

(a) 求该信道的最佳功控方案$P[i]/\overline{P}$以及相应的归一化信道容量（C/B）。

(b) 给出该信道的信道反转功控方案及相应的零中断容量（带宽归一化）。

(c) 写出该信道下的截断式信道反转功控方案，以及对应以下三个中断率的中断容量：$P_{out}=0.1$、$P_{out}=0.25$以及达到最大中断容量的P_{out}（及相关的中断门限γ_0）。

4.6 假设蜂窝系统功率衰减与距离间关系服从关系$P_r(d)=P_t(d_0/d)^{\alpha}$，其中$d_0=100$m，$\alpha$为随机变量。$\alpha$的概率分布为：$p(\alpha=2)=0.4$，$p(\alpha=2.5)=0.3$，$p(\alpha=3)=0.2$，$p(\alpha=4)=0.1$。假设接收端距离发送端$d=1000$m，发送功率上限$P_t=100$mW，接收机噪声功率0.1mW。假设收发端都已知CSI。

(a) 求接收端SNR的分布。

(b) 求该信道的最佳功控方案及相应的香农容量（C/B）。

(c) 求信道的零中断容量（带宽归一化）。

(d) 求最大中断容量（带宽归一化）。

4.7 假设有一个瑞利衰落信道，接收端和发送端都已知CSI，信噪比的概率密度函数$p(\gamma)$服从均值为$\overline{\gamma}=10$dB的指数分布。信道带宽为10MHz。

122

(a) 求使该信道容量达到香农容量的中断门限γ_0及相应的功控方案。

(b) 求该信道的香农容量。

(c) 求平均信噪比为$\overline{\gamma}=10$dB的AWGN信道的容量，并与（b）的结果比较。

(d) 求仅接收端已知CSI时的香农容量，并与（b）的结果比较。

(e) 求零中断容量及中断率为0.05的中断容量，并与（b）的结果比较。

(f) 按平均信噪比为$\overline{\gamma}=-5$dB，重新计算（b）、（c）、（d）小题。指出什么时候衰落信道的容量比AWGN大，并说明原因。

4.8 本题说明对干扰进行估计带来的容量增益，以及干扰台对系统性能的影响。考虑图4-13所示的干扰信道。信道带宽为$B=10$kHz，信号功率为10mW。信道中既有AWGN $n[k]$，也有干扰$I[k]$。噪声的功率谱密度为$N_0=10^{-8}$W/Hz。干扰$I[k]$也具有AWGN特性，干扰台接入的概率为0.25，断开的概率为0.75，干扰台功率为9mW。

(a) 假设收发端都不知道干扰台何时接入，求该情况下信道的香农容量。

(b) 假设收发端都知道干扰台何时接入，求该情况下信道的香农容量。

(c) 假设某个恶意的干扰台能够精确得知$x[k]$的内容，这样干扰源不再是AWGN。假设

接收端或发送端对干扰台的行为都一无所知，干扰台一直接入系统且平均发送功率10mW。干扰台采取什么样的策略可以使接收信噪比最小？

123

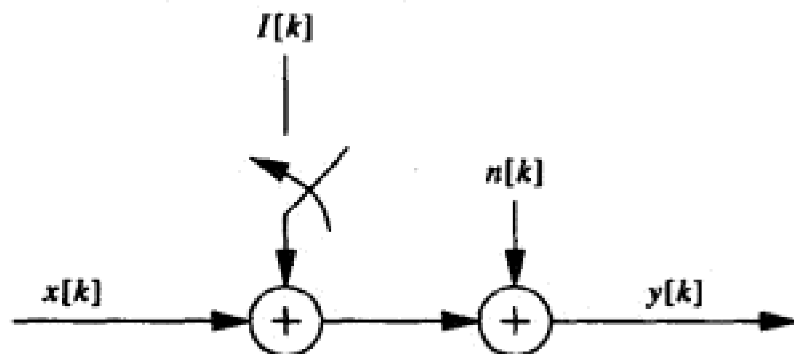


图4-13　题4.8中的干扰信道

4.9　仍考虑题4.8中的干扰台。假设发送端完全已知干扰信号。考虑以下两种传输方案：一种是发送端无视干扰信号，用其所有功率发送它的信号，另一种是发送端用一部分功率去抵消干扰（也就是发送与干扰信号极性相反的信号）。第一种情况下干扰台通过增加噪声降低了信道容量，第二种情况下干扰台通过消耗发送端的部分功率而减小了信道容量。请问哪种方案可以获得更大的信道容量？

注：还有第三种方案，它利用干扰信号的结构进行编码。这种方案称为*脏纸编码*（dirty paper coding）。如第14章所述，这种方案可用于在多天线的广播信道中逼近香农容量。

4.10　请用拉格朗日法证明：在时不变分块衰落信道中，最佳的功控方案为式（4-24）所示的注水法功控。

4.11　假设时不变分块衰落信道在$f>0$时的频率响应如下：

$$H(f)=\begin{cases}1 & f_c-20\text{ MHz}\leqslant f<f_c-10\text{ MHz}\\ 0.5 & f_c-10\text{ MHz}\leqslant f<f_c\\ 2 & f_c\leqslant f<f_c+10\text{ MHz}\\ 0.25 & f_c+10\text{ MHz}\leqslant f<f_c+20\text{ MHz}\\ 0 & \text{其他}\end{cases}$$

且频率响应满足$H(-f)=H(f)$。若平均发送功率为10mW，噪声功率谱密度为0.001μW/Hz，求最佳功控方案及香农容量。

4.12　证明：对于时不变频率选择性衰落信道，使信道容量最大的功控方案为式（4-28）所示的频域注水法功控。

4.13　某频率选择性衰落信道的总带宽是12MHz，相干带宽是$B_c=4\text{MHz}$。将该信道划分为三个带宽为B_c的子信道，并假设各子信道服从瑞利衰落且相互独立。各子信道平均增益分别为$E[|H_1[i]|^2]=1$、$E[|H_2[i]|^2]=0.5$及$E[|H_3[i]|^2]=0.125$。总发送功率为30mW，接收端噪声的功率普密度为$N_0=0.001\mu\text{W/Hz}$。

（a）假设收发端都已知瞬时$H_j[i]$，$j=1$、2、3，求该信道的最佳二维注水法功控及相应的香农容量。

（b）若将总功率平均分配在各子信道上，即每信道10mW，然后对各子信道进行时域注水法功控，求此时的信道容量并与（a）的结果作比较。

参考文献

[1] C. E. Shannon, "A mathematical theory of communication," *Bell System Tech. J.*, pp. 379–423, 623–56, 1948.
[2] C. E. Shannon, "Communications in the presence of noise." *Proc. IRE*, pp. 10–21, 1949.
[3] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*, University of Illinois Press, Urbana, 1949.
[4] M. Medard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Trans. Inform. Theory*, pp. 933–46, May 2000.
[5] R. G. Gallager, *Information Theory and Reliable Communication*, Wiley, New York, 1968.
124
[6] T. Cover and J. Thomas, *Elements of Information Theory*, Wiley, New York, 1991.
[7] C. Heegard and S. B. Wicker, *Turbo Coding*, Kluwer, Boston, 1999.
[8] I. Csiszár and J. Kórner, *Information Theory: Coding Theorems for Discrete Memoryless Channels*, Academic Press, New York, 1981.
[9] I. Csiszár and P. Narayan, "The capacity of the arbitrarily varying channel," *IEEE Trans. Inform. Theory*, pp. 18–26, January 1991.
[10] I. C. Abou-Faycal, M. D. Trott, and S. Shamai, "The capacity of discrete-time memoryless Rayleigh fading channels," *IEEE Trans. Inform. Theory*, pp. 1290–1301, May 2001.
[11] A. Lapidoth and S. M. Moser, "Capacity bounds via duality with applications to multiple-antenna systems on flat-fading channels," *IEEE Trans. Inform. Theory*, pp. 2426–67, October 2003.
[12] A. J. Goldsmith and P. P. Varaiya, "Capacity, mutual information, and coding for finite-state Markov channels," *IEEE Trans. Inform. Theory*, pp. 868–86, May 1996.
[13] M. Mushkin and I. Bar-David, "Capacity and coding for the Gilbert–Elliot channel," *IEEE Trans. Inform. Theory*, pp. 1277–90, November 1989.
[14] T. Holliday, A. Goldsmith, and P. Glynn, "Capacity of finite state Markov channels with general inputs," *Proc. IEEE Internat. Sympos. Inform. Theory*, p. 289, July 2003.
[15] R. J. McEliece and W. E. Stark, "Channels with block interference," *IEEE Trans. Inform. Theory*, pp. 44–53, January 1984.
[16] A. Lapidoth and S. Shamai, "Fading channels: How perfect need 'perfect side information' be?" *IEEE Trans. Inform. Theory*, pp. 1118–34, November 1997.
[17] G. J. Foschini, D. Chizhik, M. Gans, C. Papadias, and R. A. Valenzuela, "Analysis and performance of some basic space-time architectures," *IEEE J. Sel. Areas Commun.*, pp. 303–20, April 2003.
[18] W. L. Root and P. P. Varaiya, "Capacity of classes of Gaussian channels," *SIAM J. Appl. Math.*, pp. 1350–93, November 1968.
[19] J. Wolfowitz, *Coding Theorems of Information Theory*, 2nd ed., Springer-Verlag, New York, 1964.
[20] A. J. Goldsmith and P. P. Varaiya, "Capacity of fading channels with channel side information," *IEEE Trans. Inform. Theory*, pp. 1986–92, November 1997.
[21] G. Caire and S. Shamai, "On the capacity of some channels with channel state information," *IEEE Trans. Inform. Theory*, pp. 2007–19, September 1999.
[22] M.-S. Alouini and A. J. Goldsmith, "Capacity of Rayleigh fading channels under different adaptive transmission and diversity combining techniques," *IEEE Trans. Veh. Tech.*, pp. 1165–81, July 1999.
[23] S.-G. Chua and A. J. Goldsmith, "Variable-rate variable-power MQAM for fading channels," *IEEE Trans. Commun.*, pp. 1218–30, October 1997.
[24] S.-G. Chua and A. J. Goldsmith, "Adaptive coded modulation for fading channels," *IEEE Trans. Commun.*, pp. 595–602, May 1998.

[25] K. S. Gilhousen, I. M. Jacobs, R. Padovani, A. J. Viterbi, L. A. Weaver, Jr., and C. E. Wheatley III, "On the capacity of a cellular CDMA system," *IEEE Trans. Veh. Tech.*, pp. 303–12, May 1991.

[26] G. J. Foschini, "Layered space-time architecture for wireless communication in fading environments when using multi-element antennas," *Bell System Tech. J.*, pp. 41–59, Autumn 1996.

[27] E. Teletar, "Capacity of multi-antenna Gaussian channels," AT&T Bell Labs Internal Tech. Memo, June 1995.

[28] G. J. Foschini and M. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Pers. Commun.*, pp. 311–35, March 1998.

[29] W. Hirt and J. L. Massey, "Capacity of the discrete-time Gaussian channel with intersymbol interference," *IEEE Trans. Inform. Theory*, pp. 380–8, May 1988.

[30] A. Goldsmith and M. Medard, "Capacity of time-varying channels with channel side information," *IEEE Trans. Inform. Theory* (to appear).

[31] S. Diggavi, "Analysis of multicarrier transmission in time-varying channels," *Proc. IEEE Internat. Conf. Commun.*, pp. 1191–5, June 1997.

125



第5章　数字调制与检测

在过去的几十年中，数字信号处理技术和硬件技术的发展使数字收发器比模拟收发器更廉价、速度更快、效率更高。更为重要的是，数字调制比模拟调制有许多其他的优点，包括高频谱效率、强纠错能力、抗信道失真、高效的多址接入以及更好的安全保密性等。比如，MQAM这类多电平数字调制技术的频谱效率要比模拟调制高得多；编码技术与编码调制技术的进展使数字信号抵抗噪声和衰落的能力越来越强；而且均衡与多载波技术可以减小码间干扰（intersymbol interference，ISI）；扩频技术能消除多径或者能对多径进行合并、能抑制干扰、能检测出多个用户的传输；数字调制更易于加密，从而使数字通信有更高的安全性和保密性。正是因为这些原因，目前在建的或者将要建设的无线通信系统都是数字系统。

数字调制与解调的目的是通过信道传送比特形式的信息。比特是取值为0或1的二进制数字。信息比特可能直接来自于数字信源，也可能是模拟信源通过A/D转换器的输出。无论何种来源，送到数字调制器的信息比特有可能已经经过了压缩。数字调制将信息比特映射为模拟波形后通过信道传输，而解调则是根据信道输出的信号估计发送的比特序列。在选择具体的调制方式时，主要考虑下面几点：

- 高传输速率
- 高频带利用率（最小带宽占用）
- 高功率效率（最小发送功率）
- 对信道影响的抵抗能力（最小误比特率）
- 低功耗和低成本

126

这些要求经常是相互矛盾的，因此，调制方式的选择取决于多种因素的最佳权衡。

数字调制主要分为两类：幅度/相位调制和频率调制。频率调制用非线性方法产生，其信号包络一般是恒定的，因此称为*恒包络调制*（constant envelope modulation)或*非线性调制*（nonlinear modulation)。幅度/相位调制也称为*线性调制*（linear modulation)。线性调制一般比非线性调制有更好的频谱特性，这是因为非线性处理会导致频谱扩展。不过幅度/相位调制把信息包含在发送信号的幅度或相位中，这使它易受衰落和干扰的影响。幅度/相位调制一般需要用价格昂贵、功率效率差的线性放大器，而频率调制可以用价格便宜、功率效率高的非线性放大器。选择线性调制还是非线性调制就是在前者的频谱效率和后者的功率效率及抗信道影响能力之间进行选择。选定调制方式后，还必须要确定星座的大小。对于相同的带宽，大的星座对应高的传输速率，但大星座的调制易受噪声、衰落、硬件缺陷等的影响。另外，有些解调器需要建立一个与发送端一致的相干载波，做到这一点有一定难度，通常会大大增加接收机的复杂性。因此，不要求接收端有相干载波的调制技术更受欢迎。

本章从信号空间的概念开始讨论。将无限维的信号映射到有限维的向量空间能大大简化调制解调技术的分析与设计。接下来我们用信号空间的方法分析幅度/相位调制，包括脉幅调制（Pulse Amplitude Modulation, PAM）、相移键控（Phase-Shift Keying,

PDG

PSK）及正交幅度调制（Quadrature Amplitude Modulation，QAM）。我们还会讨论这些调制方式中的星座成形技术、正交偏移技术以及无需相干载波的差分技术。然后讨论频率调制技术及其特性，包括移频键控（Frequency Shift Keying，FSK）、最小移频键控（Minimum-Shift Keying，MSK）和连续相位移频键控（Continuous-Phase FSK，CPFSK）。本章内容还包括相干及非相干检测、能改进频谱特性的脉冲成形技术以及载波恢复及符号同步问题。

5.1 信号空间分析

数字调制将若干个比特映射为几种可能的发送信号之一。直观来说，接收机将收到的信号同各个可能的发送信号做比较，找到“最近”的那个作为检测结果，这样可以使出错的概率最小。为此我们需要一个度量来反映信号之间的距离。通过将信号投影到一组基函数上，就能将信号波形和向量表示一一对应起来。这样，问题就从无限维的函数空间转到了有限维的向量空间。这样就可以利用向量空间中现成的距离概念。本节将介绍如何把数字调制信号表示为向量，如何根据这种向量表示得到最佳的解调方法。随后的几节将用信号空间分析的方法详细分析一些具体的调制方式。

127

5.1.1 信号与系统模型

考虑图5-1所示的通信系统模型，系统每隔T秒发送$K=\log_2 M$个信息比特，数据速率为$R=K/T$bit/s。K个比特能组成2^k种不同的比特序列，每一种K比特的序列为一个消息$m_i=\{b_1,\cdots,b_k\}\in\mathcal{M}$，其中$\mathcal{M}=\{m_1,\cdots,m_M\}$是所有消息组成的集合。发送第$i$个消息的概率是$p_i$，$\sum_{i=1}^{M}p_i=1$。

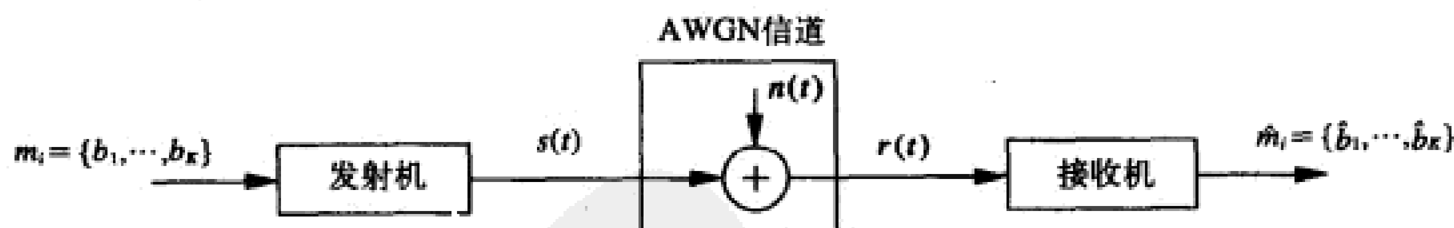


图5-1 通信系统模型

假设在$[0, T]$时间间隔内传输的消息是m_i。由于信道是模拟的，信息必须加载到适合于信道传输的模拟信号中，因此，每个消息$m_i\in\mathcal{M}$都被映射到一个特定的模拟信号$S_i(t)\in\mathcal{S}=\{s_1(t),\cdots,s_M(t)\}$上，其中$s_i(t)$定义在时间$[0, T)$上，其能量为

$$E_{s_i}=\int_0^T s_i^2(t)\,\mathrm{d}t \quad i=1,\cdots,M \tag{5-1}$$

每个消息代表一个比特序列，因此每个信号$s_i(t)\in\mathcal{S}$同样也代表一个比特序列。接收端检测出发送的$s_i(t)$等价于检测出发送的比特序列。对于发送消息构成的序列，时间区间$[kT,(k+1)T)$发送的消息m_i对应有一个模拟信号$s_i(t-kT)$，发送端发送的总信号是各时间区间内相应的模拟信号构成的序列：$s(t)=\sum_k s_i(t-kT)$。这一点如图5-2所示，图中发送的消息序列是m_1、m_2、m_1、m_1，其中消息m_i被映射为$s_i(t)$，发送的总信号为$s(t)=$

$s_1(t)+s_2(t-tT)+s_1(t-2T)+s_1(t-3T)$。

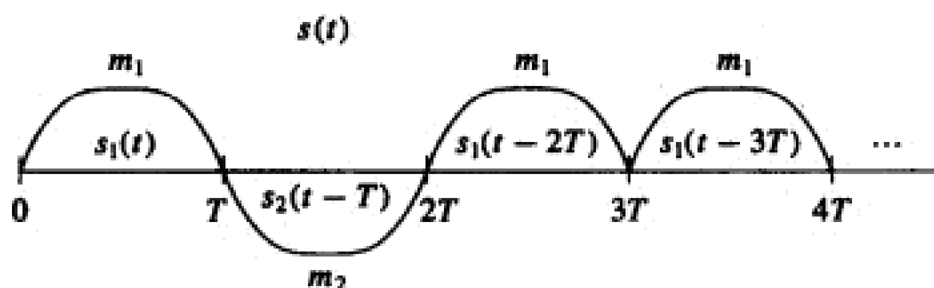


图5-2 消息序列对应的发送信号

在图5-1的模型中，信号在通过AWGN信道传输时叠加了功率谱密度为$N_0/2$的高斯白噪声，得到接收信号为$r(t)=s(t)+n(t)$。接收机将根据所收到的$r(t)$，确定出每个时间间隔$[kT, (k+1)T)$内，最有可能发送的是哪个$s_i(t)$ ($s_i(t)\in \mathcal{S}$)。对$s_i(t)$所做的最佳估计可直接映射为对消息的最佳估计，然后接收机输出最佳估计的消息$\hat{m}=\{\hat{b}_1,\cdots,\hat{b}_K\}\in \mathcal{M}$。

接收机的设计在消息估计方面的目标就是要使每个码元间隔$[kT, (k+1)T)$内估计的差错概率最小化：

128

$$P_e=\sum_{i=1}^{M} p(\hat{m}\neq m_i \mid m_i \text{ sent})p(m_i \text{ sent}) \tag{5-2}$$

把信号$\{s_i(t):i=1,\cdots, M\}$用几何方式表示，我们就可以利用最小距离准则得到AWGN信道下的最佳接收机设计。前几章中已经指出，无线信道一般是一个时变的冲激响应叠加上AWGN噪声。本书第6章将考虑这一点对数字调制性能的影响，第11章至第13章将介绍消除这些影响的方法。

5.1.2 信号的几何表示

将信号进行几何表示的基础是基的概念。通过施密特正交化[1-2]，我们可以把任意M个定义在$[0, T)$上的有限能量实信号$\mathcal{S}=\{s_1(t),\cdots, s_M(t)\}$表示为$N\leqslant M$个实正交基函数$\{\phi_1(t),\cdots, \phi_N(t)\}$的线性组合。称这组基函数张成了集合$\mathcal{S}$。每一个$s_i(t)\in \mathcal{S}$都可用这组基函数表示为

$$s_i(t)=\sum_{j=1}^{N} s_{ij}\phi_j(t) \quad 0\leqslant t<T \tag{5-3}$$

其中

$$s_{ij}=\int_0^T s_i(t)\phi_j(t)\,\mathrm{d}t \tag{5-4}$$

是将$s_i(t)$投影到基函数$\phi_j(t)$上得到的实系数。正交的含义是

$$\int_0^T \phi_i(t)\phi_j(t)\,\mathrm{d}t=\begin{cases}1 & i=j\\ 0 & i\neq j\end{cases} \tag{5-5}$$

如果信号$\{s_i(t)\}$线性无关，则$N=M$，否则$N<M$。对于持续时间为T、带宽为B的任意信号$s_i(t)$，需要的基的个数N至少是$2BT$（见文献[3]的5.3节），即这样的信号$s_i(t)$具有$2BT$维。

线性带通调制的基函数由正弦和余弦函数组成：

$$\phi_1(t)=\sqrt{\frac{2}{T}}\cos(2\pi f_c t) \tag{5-6}$$

$$\phi_2(t)=\sqrt{\frac{2}{T}}\sin(2\pi f_c t) \tag{5-7}$$

129

其中系数 $\sqrt{2/T}$ 用于归一化，即为了使$\int_0^T\phi_i^2(t)\mathrm{d}t=1$，$i=1$、2。实际上，这两个函数只是近似满足式（5-5），因为

$$\int_0^T \phi_1^2(t)\,\mathrm{d}t=\frac{2}{T}\int_0^T 0.5[1+\cos(4\pi f_c t)]\,\mathrm{d}t=1+\frac{\sin(4\pi f_c T)}{4\pi f_c T} \tag{5-8}$$

式中第二项的分子项小于1，故当$f_cT>>1$时第二项可以忽略。当$f_cT>>1$时，同样有

$$\int_0^T \phi_1(t)\phi_2(t)\,\mathrm{d}t=\frac{2}{T}\int_0^T 0.5\sin(4\pi f_c t)\,\mathrm{d}t=\frac{-\cos(4\pi f_c T)}{4\pi f_c T}\approx 0 \tag{5-9}$$

对于 $\phi_1(t)=\sqrt{2/T}\cos(2\pi f_c t)$ 和 $\phi_2(t)=\sqrt{2/T}\sin(2\pi f_c t)$ 这两个基，式（5-3）对应于带通信号的等效基带表示，两个系数s_{i1}、s_{i2}对应同相分量和正交分量：

$$s_i(t)=s_{i1}\sqrt{\frac{2}{T}}\cos(2\pi f_c t)+s_{i2}\sqrt{\frac{2}{T}}\sin(2\pi f_c t) \tag{5-10}$$

注意，作为基函数的载波可能会包含一个初始相位ϕ_0。基函数也可能会包含一个用以改善发送信号频谱特性的脉冲成形滤波器$g(t)$：

$$s_i(t)=s_{i1}g(t)\cos(2\pi f_c t)+s_{i2}g(t)\sin(2\pi f_c t) \tag{5-11}$$

此时需注意脉冲形状$g(t)$必须要保证式（5-5）的正交性，即要求

$$\int_0^T g^2(t)\cos^2(2\pi f_c t)\,\mathrm{d}t=1 \tag{5-12}$$

$$\int_0^T g^2(t)\cos(2\pi f_c t)\sin(2\pi f_c t)\,\mathrm{d}t=0 \tag{5-13}$$

其中等式在$f_cT>>1$时近似成立，如同式（5-8）、式（5-9）。若$g(t)$的带宽满足$B<<f_c$，那么$g^2(t)$在载波周期$T_c=1/f_c$内近似恒定。因为正弦余弦在一个载波周期内正交，故此式（5-13）近似成立。满足式（5-12）、式（5-13）的最简单脉冲是矩形脉冲 $g(t)=\sqrt{2/T}$，$0\leqslant t<T$。

例5.1：二元相移键控（binary phase shift keying，BPSK）调制发送1时发送$s_1(t)=\alpha\cos(2\pi f_c t)$，$0\leqslant t<T$；发送0时发送$s_2(t)=-\alpha\cos(2\pi f_c t)$，$0\leqslant t<T$。求归一化正交基函数及对应此调制的系数$\{s_{ij}\}$。

解：$s_1(t)$和$s_2(t)$对应的基函数只有一个：

$$\phi(t)=\sqrt{2/T}\cos(2\pi f_c t)$$

其中系数 $\sqrt{2/T}$是为了归一化。所求的两个系数是 $s_1=\alpha\sqrt{T/2}$和 $s_2=-\alpha\sqrt{T/2}$。

130

将表示$s_i(t)$的系数$\{s_{ij}\}$写成向量$\mathbf{s}_i=[s_{i1},\cdots,s_{iN}]^T\in\mathbb{R}^N$，称其为$s_i(t)$的*信号星座点*（signal constellation point）。所有星座点$\{\mathbf{s}_1,\cdots,\mathbf{s}_M\}$构成*信号星座图*（signal constellation）。对于给定的基函数$\{\phi_1(t),\cdots,\phi_N(t)\}$，信号$s_i(t)$与其星座点$\mathbf{s}_i$是一一对应的。由式（5-3），

可从$\mathbf{s}_i$得到$s_i(t)$；由式（5-4)，可从$s_i(t)$得到$\mathbf{s}_i$。因此，用$\mathbf{s}_i$或$s_i(t)$来表示信号是等价的。把信号$s_i(t)$用其星座点$\mathbf{s}_i \in \mathbb{R}^N$表示，就叫信号空间表示（signal space representation)，包含星座图的向量空间称为信号空间(signal space)。二维信号空间如图5-3所示，$\mathbf{s}_i \in \mathbb{R}^2$，$\mathbb{R}^2$的第$i$个坐标轴对应于基函数$\phi_i(t)$，$i=1$、2。借助于这种信号空间表示，分析无穷维的函数$s_i(t)$，就是分析$\mathbb{R}^2$上的向量$\mathbf{s}_i$。这将大大简化系统性能的分析及最佳接收机设计的推导。本章后面将给出几种常用的调制，如MPSK及MQAM等的信号空间表示，这些调制都是二维的，两个维分别对应同相基函数及正交基函数。

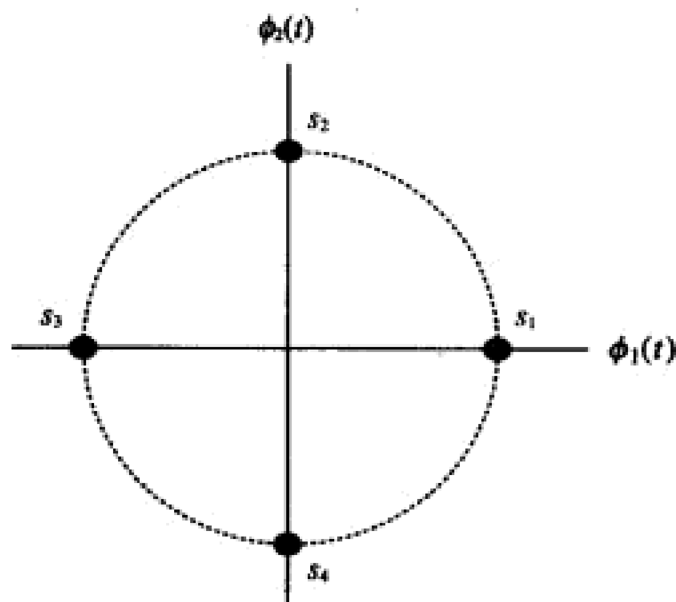


图5-3 信号空间表示

为了用信号的空间表示来分析信号，我们需要了解一些向量空间$\mathbb{R}^N$中的概念。$\mathbb{R}^N$中向量的长度定义为

$$\|\mathbf{s}_i\| \triangleq \sqrt{\sum_{j=1}^{N} s_{ij}^2} \tag{5-14}$$

两个星座点$\mathbf{s}_i$和$\mathbf{s}_k$之间的距离是

$$\|\mathbf{s}_i - \mathbf{s}_k\| = \sqrt{\sum_{j=1}^{N}(s_{ij} - s_{kj})^2} = \sqrt{\int_0^T (s_i(t) - s_k(t))^2 \, dt} \tag{5-15}$$

其中第二个等式是通过将$s_i(t)$和$s_k(t)$按式（5-3)用基函数表示，再利用基函数的正交性得到的。在$[0, T]$上，两个实信号$s_i(t)$和$s_k(t)$的内积$\langle s_i(t), s_k(t)\rangle$是

$$\langle s_i(t), s_k(t)\rangle = \int_0^T s_i(t) s_k(t) \, dt \tag{5-16}$$

类似的，两个星座点的内积$\langle \mathbf{s}_i, \mathbf{s}_k\rangle$是

$$\langle \mathbf{s}_i, \mathbf{s}_k\rangle = \mathbf{s}_i^T \mathbf{s}_k = \int_0^T s_i(t) s_k(t) \, dt = \langle s_i(t), s_k(t)\rangle \tag{5-17}$$

其中向量内积等于信号内积这一点是通过将信号按式（5-3)的基函数表示，再利用基函数的正交性得到的。若两个信号的内积为零，则称它们正交（orthogonal)。由式（5-5)可知基函数是正交函数。

131

5.1.3 接收机结构和充分统计量

已知信道输出$r(t)=s_i(t)+n(t)$，$0\leqslant t<T$，我们需要设计一个接收机结构来确定在$[0, T)$时间内发送的是哪个星座点$\mathbf{s}_i$，或者说是哪个消息m_i。在其他码元间隔内($[kT, (k+1)T)$)情

PDG

形类似。有了上面介绍的信号空间概念，我们自然会想到把接收信号$r(t)$转换为向量，以便在有限维向量空间中研究发送信号的估计问题。不过在这样做之前，首先需要确信这样的转换不会影响估计的准确性。现在我们来研究一种既能将信号转换成向量，同时又不影响性能的接收机。考虑图5-4所示的接收机结构，其中

$$s_{ij}=\int_0^T s_i(t)\phi_j(t)\,\mathrm{d}t \tag{5-18}$$

$$n_j=\int_0^T n(t)\phi_j(t)\,\mathrm{d}t \tag{5-19}$$

将$r(t)$表示为

$$\sum_{j=1}^{N}(s_{ij}+n_j)\phi_j(t)+n_r(t)=\sum_{j=1}^{N}r_j\phi_j(t)+n_r(t) \tag{5-20}$$

这里$r_j=s_{ij}+n_j$，$n_r(t)=n(t)-\sum_{j=1}^{N}n_j\phi_j(t)$代表"剩余"噪声，它是与信号空间正交的噪声分量。假如我们能够证明，给定接收信号$r(t)$时的最佳检测不会用到剩余噪声$n_r(t)$，那么最佳接收机所做出的估计$\hat{m}$就只是$\mathbf{r}=(r_1,\cdots,r_N)$的函数。如果是这样，那么$\mathbf{r}=(r_1,\cdots,r_N)$就是最佳检测$r(t)$的充分统计量。

132

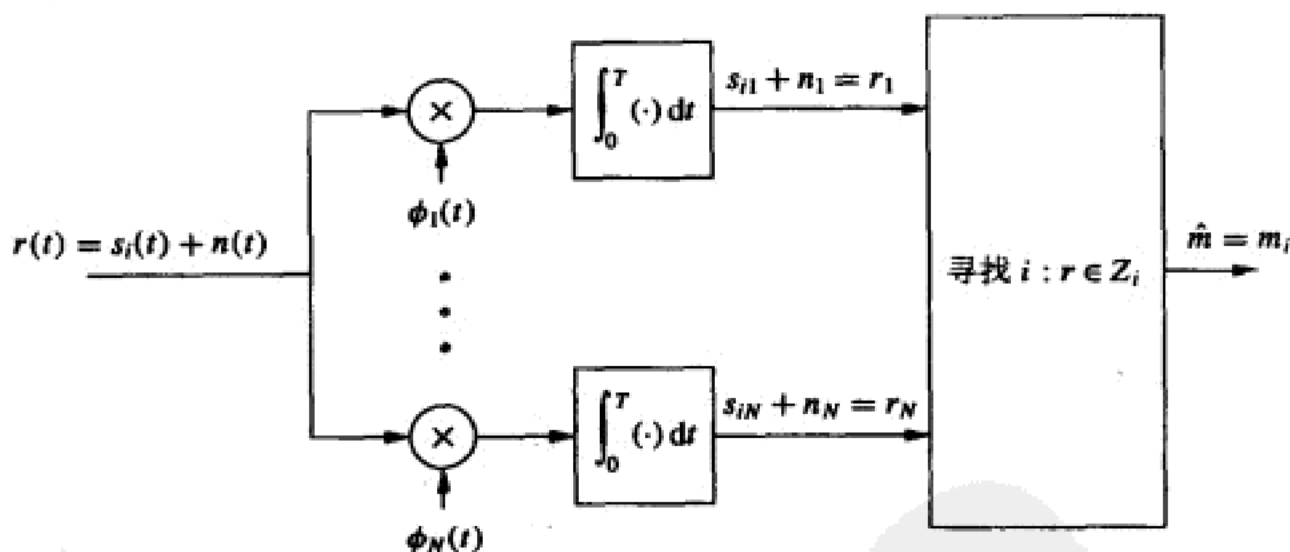


图5-4 AWGN信道下信号检测的接收机结构

从直观上来说，$n_r(t)$对信号$s_i(t)$的检测应该是没有用处的，因为它在信号空间上的投影是零。这一点如图5-5所示，图中假设发送信号处在基函数$\{\phi_1(t),\ \phi_2(t)\}$张成的空间中，剩余噪声处在由基函数$\phi_{n_r}(t)$张成的空间中，$\phi_{n_r}(t)$与$\phi_1(t)$、$\phi_2(t)$正交。把图5-5中的剩余噪声表示为$n_r$，即$n_r(t)=n_r\phi_{n_r}(t)$，则接收信号可表示为$\mathbf{r}+n_r$，其中$\mathbf{r}=(r_1,\ r_2)$，$r(t)-n_r(t)=r_1\phi_1(t)+r_2\phi_2(t)$。从图中可以看到，因为$n_r$处在和$s_i$正交的空间中，所以将$\mathbf{r}+n_r$投影到$\mathbf{r}$不会影响对$s_i$的检测。下面我们再从数学上加以论证。

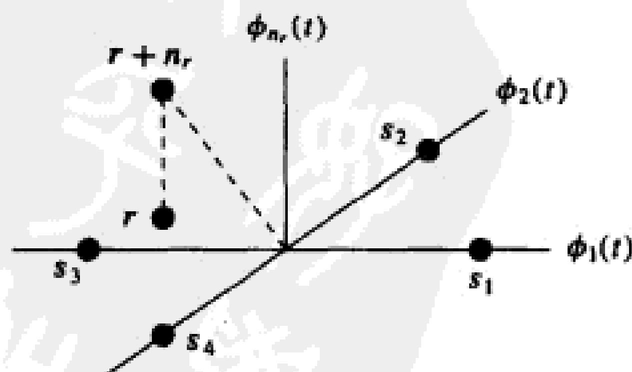


图5-5 接收信号投影为接收向量$\mathbf{r}$

首先考察$\mathbf{r}$的分布。$n(t)$是高斯过程，故在发送$s_i(t)$的条件下，信道输出$r(t)=s_i(t)+n(t)$也是高

斯过程，这样$r=(r_1,\cdots,r_N)$就是一个高斯随机向量。因为$r_j=s_{ij}+n_j$，所以在发送星座点s_i的条件下，我们有

$$\mu_{r_j|s_i}=E[r_j\mid s_i]=E[s_{ij}+n_j\mid s_{ij}]=s_{ij} \tag{5-21}$$

因$n(t)$的均值为零，故有

$$\sigma_{r_j|s_i}=E[r_j-\mu_{r_j|s_i}]^2=E[s_{ij}+n_j-s_{ij}\mid s_{ij}]^2=E[n_j^2] \tag{5-22}$$

133

$$\begin{aligned}\mathrm{Cov}[r_jr_k\mid s_i]&=E[(r_j-\mu_{r_j})(r_k-\mu_{r_k})\mid s_i]\\&=E[n_jn_k]\\&=E\left[\int_0^T n(t)\phi_j(t)\,dt\int_0^T n(\tau)\phi_k(\tau)\,d\tau\right]\\&=\int_0^T\int_0^T E[n(t)n(\tau)]\phi_j(t)\phi_k(\tau)\,dt\,d\tau\\&=\int_0^T\int_0^T \frac{N_0}{2}\delta(t-\tau)\phi_j(t)\phi_k(\tau)\,dt\,d\tau\\&=\frac{N_0}{2}\int_0^T \phi_j(t)\phi_k(t)\,dt\\&=\begin{cases}N_0/2 & j=k\\0 & j\neq k\end{cases}\end{aligned} \tag{5-23}$$

其中最后一个等式是根据基函数的正交性。因此，在发送s_i的条件下，r_j互不相关，又因为它们是高斯的，所以它们互相独立。此外，$E[n_j^2]=N_0/2$。因此r的条件分布为

$$p(r\mid s_i\ \text{sent})=\prod_{j=1}^{N}p(r_j\mid m_i)=\frac{1}{(\pi N_0)^{N/2}}\exp\left[-\frac{1}{N_0}\sum_{j=1}^{N}(r_j-s_{ij})^2\right] \tag{5-24}$$

对任意给定的$t\in[0,T)$，$n_r(t)$是高斯随机变量。容易推出$E[r_jn_r(t)|s_i]=0$，因此以s_i为条件时，r_j和$n_r(t)$互不相关，即相互独立。另外，发送的信号与噪声是相互独立的，故s_{ij}与随机过程$n_r(t)$是相互独立的。

下面我们来讨论接收机设计准则，并论证它不受$n_r(t)$的影响。接收机的设计目标是：对于给定的接收信号$r(t)$，实现对发送消息检测的错误概率最小化。即要使$P_e=p(\hat{m}\neq m_i|r(t))=1-p(\hat{m}=m_i|r(t))$最小，也即要使$p(\hat{m}=m_i|r(t))$最大。因此，收到信号$r(t)$时，接收机应该输出能使$p(\hat{m}=m_i|r(t))$最大的消息。由于消息和信号星座点之间是一对一的映射，所以也就等效于接收机应输出$p(s_i\text{sent}|r(t))$最大的消息。$r(t)$完全由$r=(r_1,\cdots,r_N)$和$n_r(t)$确定，所以有

$$\begin{aligned}p(s_i\ \text{sent}\mid r(t))&=p((s_{i1},\cdots,s_{iN})\ \text{sent}\mid(r_1,\cdots,r_N),n_r(t))\\&=\frac{p((s_{i1},\cdots,s_{iN})\ \text{sent},(r_1,\cdots,r_N),n_r(t))}{p((r_1,\ldots,r_N),n_r(t))}\\&=\frac{p((s_{i1},\cdots,s_{iN})\ \text{sent},(r_1,\cdots,r_N))p(n_r(t))}{p(r_1,\ldots,r_N)p(n_r(t))}\\&=p((s_{i1},\cdots,s_{iN})\ \text{sent}\mid(r_1,\cdots,r_N))\end{aligned} \tag{5-25}$$

其中第三个等式成立是因为$n_r(t)$和$r=(r_1,\cdots,r_N)$、$(s_{i1},\cdots,s_{iN})$都独立。这个结果不包含

$n_r(t)$，只包括向量$\boldsymbol{r}$，就是说在检测发送消息时，弃掉剩余噪声，只用$\boldsymbol{r}$进行检测就可以使错判概率最小。也就是说，$\boldsymbol{r}=(r_1,\cdots,r_N)$是接收信号$r(t)$的充分统计。我们称$\boldsymbol{r}$是与$r(t)$对应的接收向量。

5.1.4 判决域及最大似然判决准则

从前面的分析中可以看出，最佳接收机通过选择$\hat{m}$使$1-P_e=p(\hat{m}\text{sent}|\boldsymbol{r})$)最大而使错误率最小。也就是说，对于给定的接收向量$\boldsymbol{r}$，最佳接收机所选择的$\hat{m}=m_i$对应的星座点s_i满足$p(s_i\text{sent}|\boldsymbol{r})\geq p(s_j\text{sent}|\boldsymbol{r})$，$\forall j\neq i$。定义判决域（decision region）$Z_1,\cdots,Z_M$为信号空间$\mathbb{R}^N$的子集

$$Z_i=\{\boldsymbol{r}: p(s_i \text{ sent} \mid \boldsymbol{r})>p(s_j \text{ sent} \mid \boldsymbol{r})\ \forall j\neq i\} \tag{5-26}$$

134

$Z_1,\cdots,Z_M$显然互不重叠。若不存在这样的点$\boldsymbol{r}\in\mathbb{R}^N$使得$p(s_i\text{sent}|\boldsymbol{r})=p(s_j\text{sent}|\boldsymbol{r})$，则这组判决域$Z_1,\cdots,Z_M$构成了$R^N$的一个判决域。如果存在这种概率相等的点，可将这些点任意划归到Z_i或Z_j，使$Z_1,\cdots Z_M$仍然构成一个判决域。信号空间被划分为许多判决域之后，若接收向量$\boldsymbol{r}\in Z_i$，则最佳接收机的判决输出是$\hat{m}=m_i$。于是，接收机的工作是：由$r(t)$计算接收向量$\boldsymbol{r}$，找到$\boldsymbol{r}$所在的判决域Z_i，输出对应的消息m_i。图5-6中，一个二维信号空间划分成四个判决域$Z_1,\cdots,Z_4$，与之对应的星座点是$s_1,\cdots,s_4$。接收向量$\boldsymbol{r}$处在区域Z_1中，所以接收机输出消息m_1作为对接收向量$\boldsymbol{r}$的最佳估计。

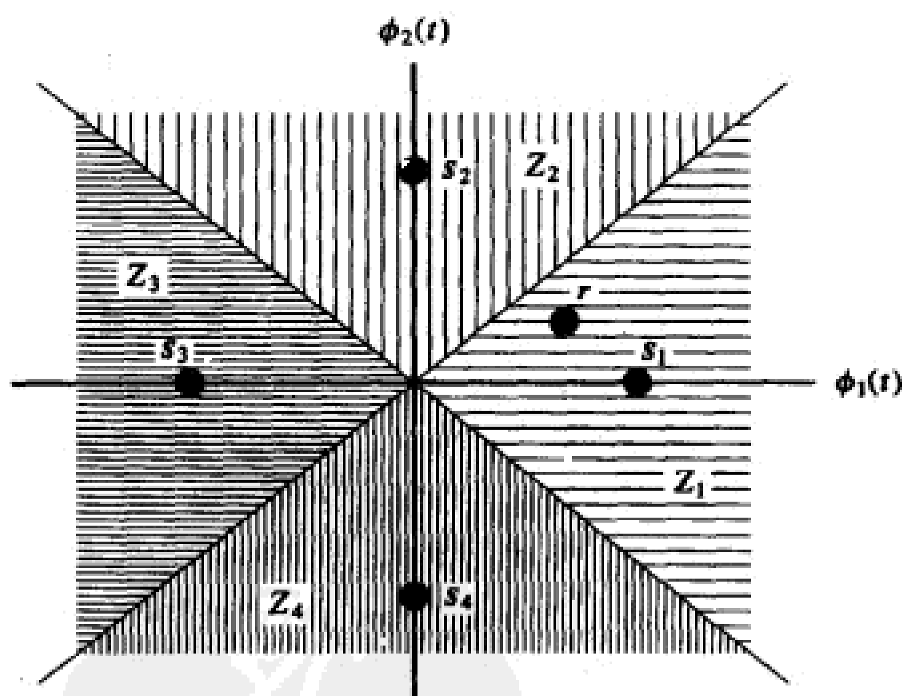


图5-6 判决域

下面我们将详细讨论判决域将$p(s_i\text{ send}|\boldsymbol{r})$简记为$p(s_i|\boldsymbol{r})$，将$p(s_i\text{ sent})$简记为$p(s_i)$。由贝叶斯公式

$$p(s_i \mid \boldsymbol{r})=\frac{p(\boldsymbol{r} \mid s_i)p(s_i)}{p(\boldsymbol{r})} \tag{5-27}$$

为了使错误率最小，接收机应输出能使$p(s_i|\boldsymbol{r})$最大的s_i所对应的$\hat{m}=m_i$，即s_i必须满足

$$\arg\max_{s_i}\frac{p(\boldsymbol{r} \mid s_i)p(s_i)}{p(\boldsymbol{r})}=\arg\max_{s_i} p(\boldsymbol{r} \mid s_i)p(s_i) \quad i=1,\cdots,M \tag{5-28}$$

等式成立是因为$p(\boldsymbol{r})$与s_i无关。设消息等概，即$p(s_i)=1/M$，接收机输出的$\hat{m}=m_i$所对应的s_i须满足

$$\arg\max_{s_i} p(\boldsymbol{r} \mid s_i) \quad i=1,\cdots,M \tag{5-29}$$

定义似然函数为

135

$$L(s_i)=p(\boldsymbol{r} \mid s_i) \tag{5-30}$$

对于给定的接收向量$\boldsymbol{r}$，最大似然接收机（maximum likelihood receiver）输出的$\hat{m}=m_i$对应于能使似然函数$L(s_i)$最大的s_i。由于对数函数是增函数，所以使$L(s_i)$最大相当于使其对数最大。另外，式（5-24）中的常数$(\pi N_0)^{-N/2}$并不影响使$L(s_i)$最大。这样，最大化$L(s_i)$等价于最大化对数似然函数（log likelihood function）$l(s_i)=\ln[(\pi N_0)^{-N/2}L(s_i)]$，将式（5-24）代入式（5-30）得

$$l(s_i)=-\frac{1}{N_0}\sum_{j=1}^{N}(r_j-s_{ij}^2)=-\frac{1}{N_0}\|\boldsymbol{r}-s_i\|^2 \tag{5-31}$$

即$l(s_i)$只取决于接收向量$\boldsymbol{r}$与星座点s_i之间的距离。从式（5-31）进一步可知，与接收向量$\boldsymbol{r}$距离最近的s_i能使$l(s_i)$最大。

最大似然接收机的实现如图5-4所示。首先由$r(t)$计算出$\boldsymbol{r}$，然后从星座图中找出与$\boldsymbol{r}$最近的星座点s_i，即

$$\arg\min_{s_i}\sum_{j=1}^{N}(r_j-s_{ij})^2=\arg\min_{s_i}\|\boldsymbol{r}-s_i\|^2 \tag{5-32}$$

此s_i是由包含$\boldsymbol{r}$的判决域Z_i所决定的，Z_i由下式确定

$$Z_i=\{\boldsymbol{r}:\|\boldsymbol{r}-s_i\|<\|\boldsymbol{r}-s_j\|\ \forall j=1,\cdots,M,\ j\neq i\} \quad i=1,\cdots,M \tag{5-33}$$

最后由所估计的星座点s_i映射为接收机输出的消息$\hat{m}$。最大似然接收机认为发送的星座点是离接收向量最近的那个，这样的判决方法很符合直观感觉。最大似然接收机的判决只跟向量之间的距离有关，所以结构很简单。当发送的消息等概出现时，这种结构的接收机可使错判概率最小。但若发送消息并非等概出现，那么最大似然接收机不能使错判概率最小。为了使错判概率最小，需根据消息的概率分布按照式（5-27）对判决域进行修正。

136

图5-7所示为另一种接收机结构。它采用一组与基函数匹配的滤波器。我们称冲激响应为$\psi(t)=\phi(T-t)$, $0\leqslant t\leqslant T$的滤波器为信号$\phi(t)$的匹配滤波器，因此图5-7所示的也叫做匹配滤波接收机。可以证明，信号通过与其匹配的滤波器后，输出信噪比可达到最大。还可以证明，图5-7中匹配滤波器输出的抽样$(r_1,\cdots,r_n)$与图5-4中的$(r_1,\cdots,r_n)$完全一样，所以这两种接收机是等价的。

例5.2：对于BPSK信号，求对应星座点$s_1=A$，$s_2=-A$（$A>0$）的判决域Z_1和Z_2。

解：信号空间是一维的，因此$\boldsymbol{r}=r\in\mathbb{R}$。由式（5-33），判决域$Z_1\subset\mathbb{R}$定义为

$$Z_1=\{r:\|r-A\|<\|r-(-A)\|\}=\{r:r>0\}$$

因此Z_1包括实轴上的所有正数。类似地

$$Z_2=\{r:\|r-(-A)\|<\|r-A\|\}=\{r:r<0\}$$

因此Z_2包括实轴上的所有负数。当$r=0$时，它与s_1和s_2的距离相等，我们可以任意将其划入Z_1或Z_2。

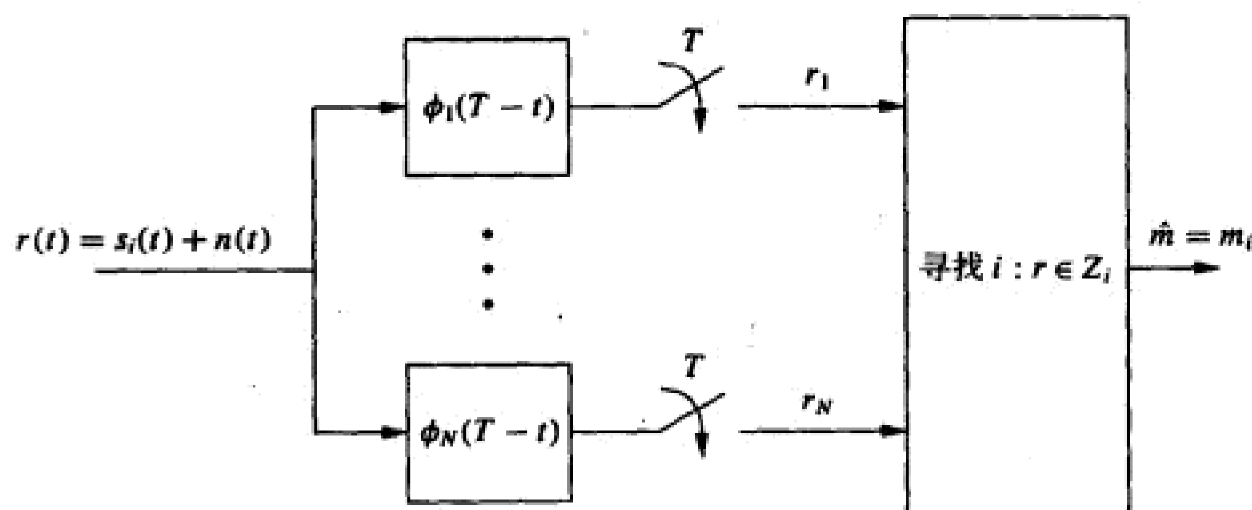


图5-7 匹配滤波接收机结构

5.1.5 误码率及联合界

接下来我们讨论误码率以及极大似然接收机的结构对于等概消息$p(m_i, \text{sent})=1/M$，有

$$\begin{aligned}
P_e &= \sum_{i=1}^{M} p(\boldsymbol{r}\notin Z_i \mid m_i\ \text{sent})p(m_i\ \text{sent}) \\
&= \frac{1}{M}\sum_{i=1}^{M} p(\boldsymbol{r}\notin Z_i \mid m_i\ \text{sent}) \\
&= 1-\frac{1}{M}\sum_{i=1}^{M} p(\boldsymbol{r}\in Z_i \mid m_i\ \text{sent}) \\
&= 1-\frac{1}{M}\sum_{i=1}^{M}\int_{Z_i} p(\boldsymbol{r}\mid m_i)\,\mathrm{d}\boldsymbol{r} \\
&= 1-\frac{1}{M}\sum_{i=1}^{M}\int_{Z_i} p(\boldsymbol{r}=\boldsymbol{s}_i+\boldsymbol{n}\mid \boldsymbol{s}_i)\,\mathrm{d}\boldsymbol{n} \\
&= 1-\frac{1}{M}\sum_{i=1}^{M}\int_{Z_i-\boldsymbol{s}_i} p(\boldsymbol{n})\,\mathrm{d}\boldsymbol{n}
\end{aligned} \tag{5-34}$$

式中的积分范围是N维空间$\mathbb{R}^N$的子集Z_i。考虑图5-8，$S_1,\cdots,S_8$以最小相互距离$d_{\min}$均分圆,积分范围Z_1如图中阴影区所示。发送$\boldsymbol{s}_1$时正确接收的概率为$p(\boldsymbol{r}\in Z_1|\boldsymbol{s}_1)$，等于把噪声加到星座点$\boldsymbol{s}_1$上之后的结果向量$\boldsymbol{r}=\boldsymbol{s}_1+\boldsymbol{n}$还在判决域$Z_1$中的概率$p(\boldsymbol{r}=\boldsymbol{s}_1+\boldsymbol{n}|\boldsymbol{s}_1)$。 137

从图5-8容易看出，误码率积分不随星座图的旋转和平移而改变。图中右侧的星座图是左侧星座图旋转$\boldsymbol{\theta}$，再平移$\boldsymbol{P}$得到的，即$s_i'=s_ie^{j\theta}+\boldsymbol{P}$。噪声向量$\boldsymbol{n}=(n_1,\cdots,n_N)$的元素是零均值、独立同分布的高斯随机变量，其极坐标表示为$\boldsymbol{n}=|\boldsymbol{n}|e^{j\theta}$，因为$\theta$服从均匀分布，

所以噪声不受相位旋转的影响，即误码率积分有旋转不变性。平移不变性是因为：如果坐标平移了$P\in\mathbb{R}^N$，则由式（5-33）定义的判决域也会平移相同的量。令(s_i, Z_i)表示平移前的星座点及其判决域，(s_i', Z_i')表示平移后的星座点及其判决域。明显有$p(r=s_i+n\in Z_i|s_i)=p(r'=s_i'+n\in Z_i'|s_i')$。因此，误码率在星座图轴平移后保持不变。

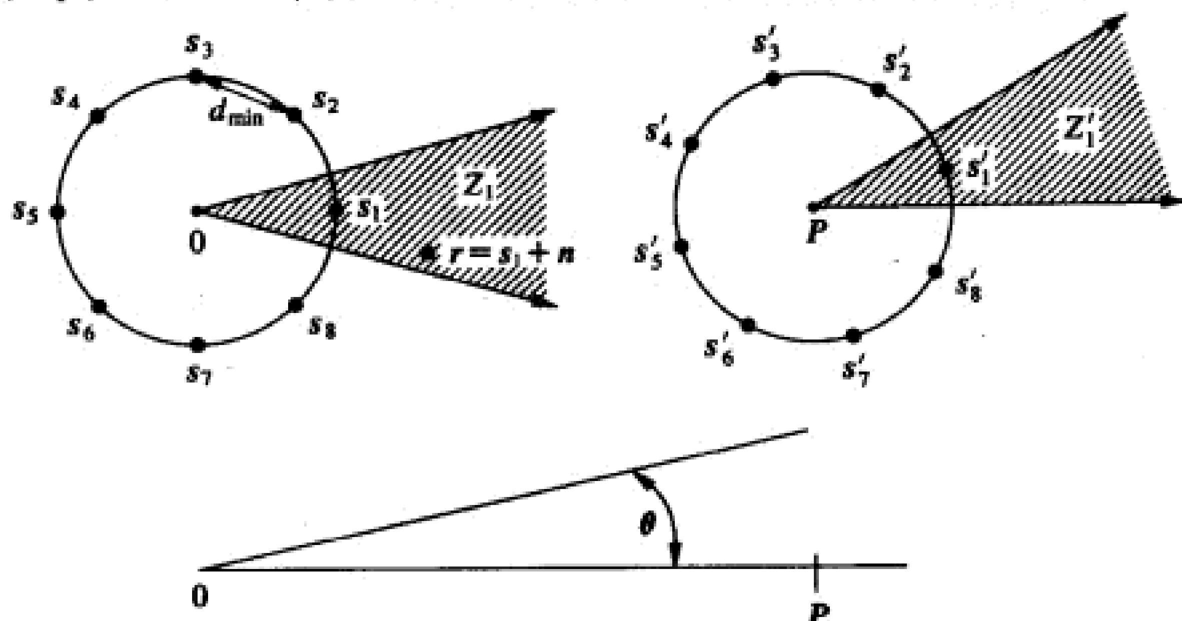


图5-8 错误率积分及其旋转/平移不变性

虽然式（5-34）给出了错误率的精确解，但一般不能得到式中积分的闭式结果。为此，我们来考虑错误率的联合界，它能写出一个用星座点间距离表示的闭式结果。设事件A_{ik}表示发送信号星座点为s_i时$\|r-s_k\|<\|r-s_i\|$。如果事件A_{ik}发生，则发送的星座点并非离接收向量最近的，判决将出现错误。不过事件A_{ik}并不必然意味着接收机将判为s_k，因为可能有另一个星座点s_j满足$\|r-s_j\|<\|r-s_k\|<\|r-s_i\|$。若对所有的$k\neq i$，有$\|r-s_i\|<\|r-s_k\|$，即当所有$A_{ik}$, $k\neq i$都不发生时，判决才是正确的。因此

$$P_e(m_i \text{ sent}) = p\left(\bigcup_{\substack{k=1\\k\neq i}}^{M} A_{ik}\right) \leqslant \sum_{\substack{k=1\\k\neq i}}^{M} p(A_{ik}) \tag{5-35}$$

138 其中的不等式就是概率论中的联合界。

对于$p(A_{ik})$，我们有

$$\begin{aligned} p(A_{ik}) &= p(\|s_k - r\| < \|s_i - r\| \mid s_i \text{ sent}) \\ &= p(\|s_k - (s_i + n)\| < \|s_i - (s_i + n)\|) \\ &= p(\|n + s_i - s_k\| < \|n\|) \end{aligned} \tag{5-36}$$

就是说，错判概率等于噪声n离矢量s_k-s_i比离原点更近的概率。噪声的均值为零，所以通常离原点较近。式（5-36）这个概率只与噪声在矢量s_k-s_i上的投影有关，如图5-9所示。由n的特性可知，它在一维直线上的投影是一个均值为零、方差为$N_0/2$的高斯随机变量n。当n离s_k-s_i比离原点更近时，事件A_{ik}发生，此即$n>d_{ik}/2$，$d_{ik}=\|s_i-s_k\|$是这两个星座点间的距离。因此

$$p(A_{ik}) = p\left(n > \frac{d_{ik}}{2}\right) = \int_{d_{ik}/2}^{\infty} \frac{1}{\sqrt{\pi N_0}} \exp\left[\frac{-v^2}{N_0}\right] \mathrm{d}v = Q\left(\frac{d_{ik}}{\sqrt{2N_0}}\right) \tag{5-37}$$

将式（5-37）代入式（5-35）得

$$P_e(m_i \text{ sent}) \leqslant \sum_{\substack{k=1\\k\neq i}}^{M} Q\left(\frac{d_{ik}}{\sqrt{2N_0}}\right) \tag{5-38}$$

其中的Q函数$Q(z)$定义为均值为零，方差为1的高斯随机变量X大于定值z的概率：

$$Q(z) = p(X > z) = \int_z^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \tag{5-39}$$

将式（5-38）对所有可能的消息求和得到联合界为

$$P_e = \sum_{i=1}^{M} p(m_i) P_e(m_i \text{ sent}) \leqslant \frac{1}{M} \sum_{i=1}^{M} \sum_{\substack{k=1\\k\neq i}}^{M} Q\left(\frac{d_{ik}}{\sqrt{2N_0}}\right) \tag{5-40}$$

注意Q函数无闭式解。它与互补误差函数有如下关系

$$Q(z) = \frac{1}{2} \text{erfc}\left(\frac{z}{\sqrt{2}}\right) \tag{5-41}$$

Q函数有如下的闭式上界

$$Q(z) \leqslant \frac{1}{z\sqrt{2\pi}} e^{-z^2/2} \tag{5-42}$$

这个上界当$z>>0$时是比较紧的。

139

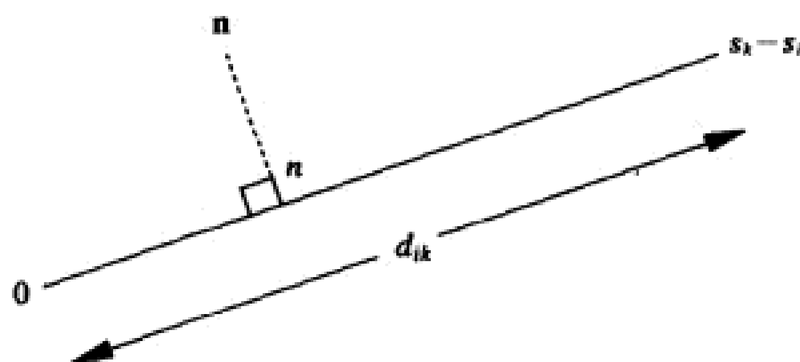


图5-9 噪声投影

定义星座图上的最小距离为$d_{\min} = \min_{i,k} d_{ik}$，则式（5-40)可简化为一个较松的界

$$P_e \leqslant (M-1) Q\left(\frac{d_{\min}}{\sqrt{2N_0}}\right) \tag{5-43}$$

代入式（5-42)，得到闭式界

$$P_e \leqslant \frac{M-1}{d_{\min}\sqrt{\pi/N_0}} \exp\left[\frac{-d_{\min}^2}{4N_0}\right] \tag{5-44}$$

有时候也这样近似，即用星座图上最小距离$d_{\min}$对应的误码率乘以距离为$d_{\min}$的邻近星座点的个数$M_{d_{\min}}$：

$$P_e \approx M_{d_{\min}} Q\left(\frac{d_{\min}}{\sqrt{2N_0}}\right) \tag{5-45}$$

称此近似为P_e的最近邻近似。当不同星座点的最近邻个数或最近邻距离不相同时，可按

各个星座点进行平均。因为$M \geqslant M_{d_{\min}}$，所以最近邻近似总小于式（5-43）的松散界。它也比式（5-40）的联合界略小，因为没有包括距离大于$d_{\min}$的星座点。不过在高信噪比时，最近邻近似非常接近误码率的准确值，这是因为高信噪比时，对于$x>y$有$Q(x)<<Q(y)$。就是说，在高信噪比时，将星座点错判为与之不相邻的星座点的概率可以忽略。式（5-45）的严格推导见文献[4~5]。文献[4]中还指出，通过适当调整系数，式（5-45）还能反映载频漂移等接收机的不理想因素引起的误码性能恶化。最近邻近似的优点是它只取决于星座图中的最小距离以及相邻星座点的个数。

例5.3：某二维星座图的四个星座点是$s_1=(A, 0)$、$s_2=(0, A)$、$s_3=(-A, 0)$、$s_4=(0, -A)$。设$A/\sqrt{N_0}=4$，试求出该星座图的最小距离，并求P_e的联合界（式（5-40））、松散界（式（5-43））、闭式界（式（5-44））及最近邻近似（式（5-45））。

140

解：此星座图如图5-3所示，圆的半径为A。由于对称性，只需考虑其中一个星座点的错误率，我们考虑s_1。易知星座点间的最小距离为$d_{\min}=d_{12}=d_{23}=d_{34}=d_{14}=\sqrt{2A^2}$。其他距离是$d_{13}=d_{24}=2A$。由对称性，对于$j \neq i$有$P_e(m_i \text{ sent})=P_e(m_j \text{ sent})$，这样联合界简化为

$$
\begin{aligned}
P_e &\leqslant \sum_{j=2}^{4} Q\left(\frac{d_{1j}}{\sqrt{2N_0}}\right) \\
&= 2Q\left(\frac{A}{\sqrt{N_0}}\right) + Q\left(\frac{\sqrt{2}A}{\sqrt{N_0}}\right) = 2Q(4) + Q(\sqrt{32}) = 3.1679\times10^{-5}
\end{aligned}
$$

松散界为

$$P_e \leqslant 3Q(4) = 9.5014\times10^{-5}$$

近似为联合界的3倍。闭式界为

$$P_e \leqslant \frac{3}{\sqrt{2\pi A^2/N_0}}\exp\left[\frac{-0.5A^2}{N_0}\right] = 1.004\times10^{-4}$$

它比联合界高一个数量级。最近邻近似为

$$P_e \approx 2Q(4) = 3.1671\times10^{-5}$$

与联合界近似相等。

注意对于二元调制（$M=2$），只有一种错判方式，$d_{\min}$是这两个星座点间的距离，式（5-43）的界变成精确值：

$$P_b = Q\left(\frac{d_{\min}}{\sqrt{2N_0}}\right) \tag{5-46}$$

在第6章中我们将看到，式（5-44）和式（5-46）中的$d_{\min}$的平方与接收信噪比成正比，因此提高接收信号功率将使误码率降低。

注意P_e是误码元率（误消息率）：$P_e=p(\hat{m} \neq m_i | m_i)$，每个码元或消息包含有$\log_2 M$比特。系统设计者一般更关心误比特率（bit error rate, BER）而不是误码率，因为误比特率直接影响上层网络协议及端到端的性能。由于邻近判决域的错误最容易发生，所以我们希望能设计出一种将M种比特组合映射到消息$m_i(i=1,\cdots, M)$，它能使邻近判决域的错误发生时，只出现1比特错误。若非邻近星座点的错误概率非常小，则这种映射下的误

比特率近似为

$$P_b \approx \frac{P_e}{\log_2 M} \tag{5-47}$$

最常见的这种将星座点误判为邻近星座点导致单一比特错误的映射是格雷码，5.3节将对此进行详细介绍。

本节所讨论的信号空间的概念适用于任何将比特组映射为模拟波形的调制，包括下面将要介绍的幅度/相位调制及频率调制。

141

5.2 带通调制原理

数字带通调制的基本原理是用载波信号携带信息比特流在信道中传输，在接收端，解调器从接收到的信号中提取出这些信息比特流。信道对发送信号的损伤可能会导致解调出现比特错误。调制的目标就是以最小的数据损伤概率实现高速传输数据。

一般来说，已调信号通过幅度$\alpha(t)$、频率$f(t)$或者相位$\theta(t)$来携带信息，可以表示为

$$s(t)=\alpha(t)\cos[2\pi(f_c+f(t))t+\theta(t)+\phi_0]=\alpha(t)\cos(2\pi f_c t+\phi(t)+\phi_0) \tag{5-48}$$

其中$\phi(t)=2\pi f(t)t+\theta(t)$，$\phi_0$是载波初相。在这个表达式中，相位和频率调制统一体现为角度调制。

将式（5-48）的右侧写为同相分量和正交分量的形式

$$\begin{aligned} s(t)&=\alpha(t)\cos(\phi(t)+\phi_0)\cos(2\pi f_c t)-\alpha(t)\sin(\phi(t)+\phi_0)\sin(2\pi f_c t)\\ &=s_I(t)\cos(2\pi f_c t)-s_Q(t)\sin(2\pi f_c t)\end{aligned} \tag{5-49}$$

其中$s_I(t)=\alpha(t)\cos(\phi(t)+\phi_0)$是$s(t)$的同相分量，$S_Q(t)=\alpha(t)\sin(\phi(t)+\phi_0)$是$s(t)$的正交分量。也可以将$s(t)$用其等效基带信号表示

$$s(t)=\mathrm{Re}\{u(t)\mathrm{e}^{\mathrm{j}2\pi f_c t}\} \tag{5-50}$$

其中$u(t)=s_I(t)+\mathrm{j}s_Q(t)$。接收机通常是分别对同相和正交分量进行基带处理，所以这种表示非常方便。附录A中对此有更详细的说明。

5.3 幅度/相位调制

幅度/相位调制是将信息调制到幅度和相位中。在码元间隔$0\leqslant t<T_s$内，信号$s(t)$的幅度或相位携带了$K=\log_2 M$个比特。可将信号$s(t)=s_I(t)\cos(2\pi f_c t)-s_Q(t)\sin(2\pi f_c t)$写成信号空间的形式为$s(t)=s_{i1}\phi_1(t)+s_{i2}\phi_2(t)$，其中基函数$\phi_1(t)=g(t)\cos(2\pi f_c t+\phi_0)$、$\phi_2(t)=-g(t)\sin(2\pi f_c t+\phi_0)$。$g(t)$是成形脉冲。设$\phi_0=0$，在$[kT,(k+1)T)$时间内传送第$i$个消息时，我们取$s_I(t)=s_{i1}g(t-KT)$、$s_Q(t)=s_{i2}g(t-KT)$。同相分量和正交分量都是基带信号，其频谱特性由成形脉冲$g(t)$决定，它们的带宽B与$g(t)$的带宽相等。发送信号$s(t)$是中心频率为f_c的带通信号，带宽是$2B$。令$B=K_g/T_s$，其中K_g由脉冲形状决定：对于矩形脉冲$K_g=1$[1]，对于升余弦脉冲，$0.5\leqslant K_g\leqslant 1$，这些将在5.5节介绍。采用矩形脉冲时$g(t)$的带宽是$1/T_s$[1]，

1. 原书为$K_g=0.5$，疑有误，此处已改。——译者注

$s(t)$的带宽是$2/T_s$[2]。幅度/相位调制的星座图由星座点$\{(s_{i1}, s_{i2})\in \mathbb{R}^2, i=1,\cdots, M\}$决定。$S(t)$的等效基带表示为

142

$$s(t) = \text{Re}\{x(t)e^{j\phi_0}e^{j2\pi f_c t}\} \tag{5-51}$$

其中$x(t)=(s_{i1}+js_{i2})g(t)$。星座点$s_i=(s_{i1}, s_{i2})$是$\log_2 M$个比特所对应的符号，$T_s$是符号间隔或码元间隔。比特率是$K$比特/符号或$R=\log_2 M/T_s$比特/秒。

幅度/相位调制主要分3种：

- 脉冲幅度调制（MPAM）：只有幅度携带信息；
- 相移键控（MPSK）：只有相位携带信息；
- 正交幅度调制(MQAM)：幅度和相位都携带信息。

数字调制的设计由每符号携带的信息量$K=\log_2 M$、信号星座$\{s_i, i=1,\cdots, M\}$以及脉冲成形函数$g(t)$决定。设计适当的脉冲可改善频谱效率并减少符号间干扰，这些将在5.5节介绍。

143

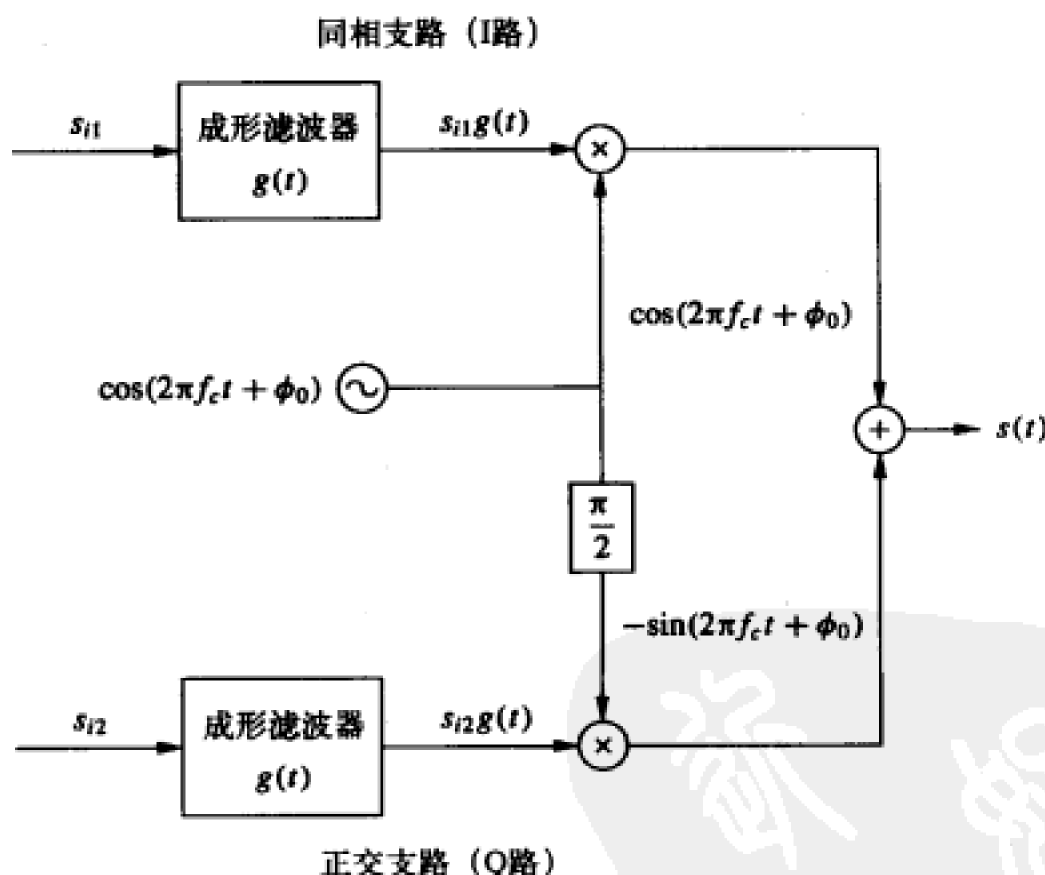


图5-10　幅度/相位调制器

幅度/相位调制可以用图5-10所示的调制器实现。基函数的相位ϕ_0是发送振荡器产生的一个任意值。解调器的结构如图5-11所示，它与图5-7等价，对应$\phi_1(t)=g(t)\cos(2\pi f_c t+\phi)$, $\phi_2(t)=-g(t)\sin(2\pi f_c t+\phi)$。接收机一般设有一个载波相位恢复电路，使接收机相位ϕ与发送端相位ϕ_0一致[3]，这种方式叫相干检测（coherent detection）。如果$\phi-\phi_0$

1. 原书为$0.5/T_s$，疑有误，此处已改。——译者注
2. 原书为$1/T_s$，疑有误，此处已改。——译者注
3. 实际当中，信道中的传播时延τ也会引起一个相位移$-2\pi f_c\tau$，因此相干检测要求接收相位为$\phi=\phi_0-2\pi f_c\tau$，详见5.6节。

$=\Delta\phi\neq 0$，那么I路中就会包含一个不应该有的正交分量，Q路中也会包含一个同相分量，即$r_1=s_{i1}\cos(\Delta\phi)+s_{i2}\sin(\Delta\phi)+n_1$、$r_2=s_{i1}\sin(\Delta\phi)+s_{i2}\cos(\Delta\phi)+n_2$，这种I、Q干扰将导致严重的性能下降。此外，还需假设图中的采样与发送码元时间同步，这一点是由定时提取电路完成的。定时和载波的相位恢复是接收机中比较复杂的一部分，在无线环境下有相当大的难度，详见5.6节。以下我们在MPAM、MPSK、MQAM中假设载波都是同步的，即$\phi=\phi_0=0$。

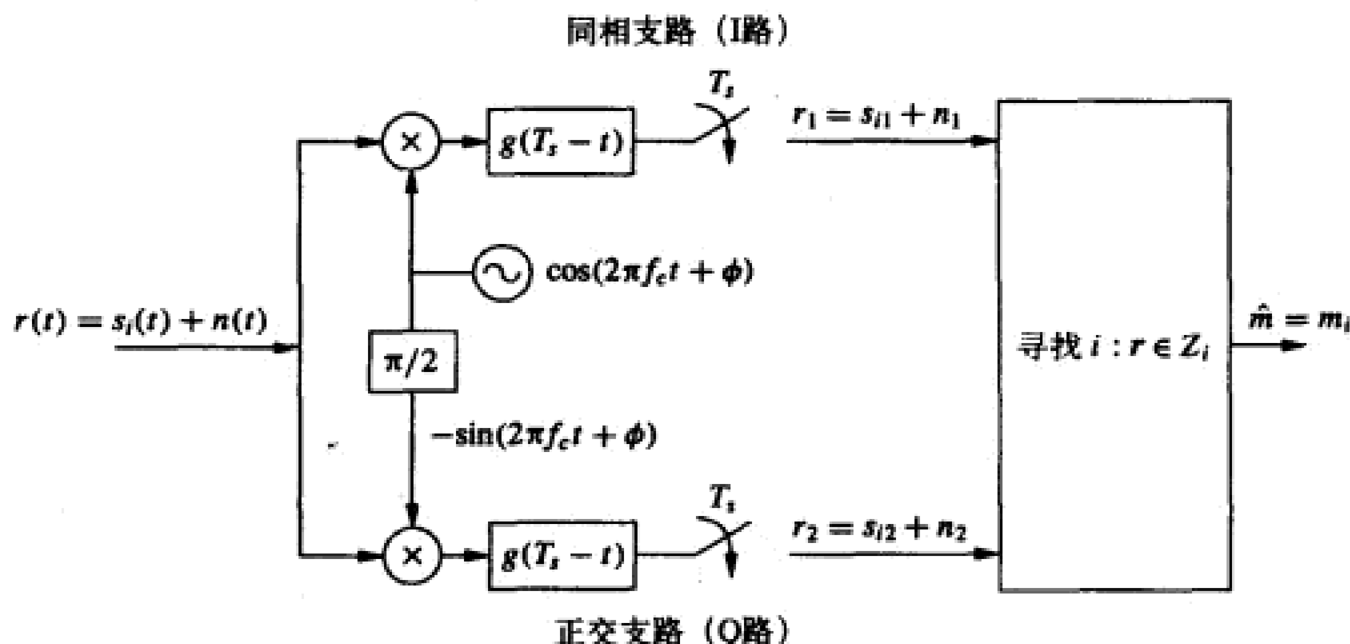


图5-11 幅度/相位解调器(相干：$\phi=\phi_0$)

5.3.1 脉冲幅度调制(MPAM)

一维的MPAM是最简单的线性调制，它没有正交分量。MPAM调制用幅度A_i携带信息，发送信号为

$$s_i(t)=\mathrm{Re}\{A_i g(t)\mathrm{e}^{\mathrm{j}2\pi f_c t}\}=A_i g(t)\cos(2\pi f_c t),\quad 0\leqslant t\leqslant T_s\gg 1/f_c \tag{5-52}$$

其中$A_i=(2i-1-M)d$, $i=1, 2,\cdots, M$。星座图$\{A_i, i=1,\cdots, M\}$由参数d决定，d一般由信号能量决定。$g(t)$是满足式（5-12）和式（5-13）的成形脉冲。星座图上的最小距离是$d_{\min}=\min_{i,j}|A_i-A_j|=2d$。发送端信号的幅度有$M$个可能值，在每个符号间隔时间内，一个码元携带了$K=\log_2 M$比特的信息。

144

由于成形脉冲满足式（5-12）[1]，所以每个码元周期内第i个星座点对应的MPAM信号的能量是

$$E_{s_i}=\int_0^{T_s}s_i^2(t)\,\mathrm{d}t=\int_0^{T_s}A_i^2 g^2(t)\cos^2(2\pi f_c t)\,\mathrm{d}t=A_i^2 \tag{5-53}$$

对于不同的$s_i(t)$，$i=1,\cdots,M$，它们的能量是不同的。假设码元等概，则平均能量为

$$\bar{E}_s=\frac{1}{M}\sum_{i=1}^{M}A_i^2 \tag{5-54}$$

1. 式（5-53）和式（5-12）一样，不是精确等式，但当f_cT_s>>1时近似程度非常好。

MPAM的星座映射通常采用如图5-12所示的格雷码映射，相邻星座点只有一比特不同。由于最容易出现的判决错误是将发送的星座点判决为邻近的星座点，因此在格雷码映射时，K个比特中只错一个比特。MPSK和矩形[1]星座的MQAM都可以设计成格雷码映射，但某些矩形星座MQAM不能实现这一点。

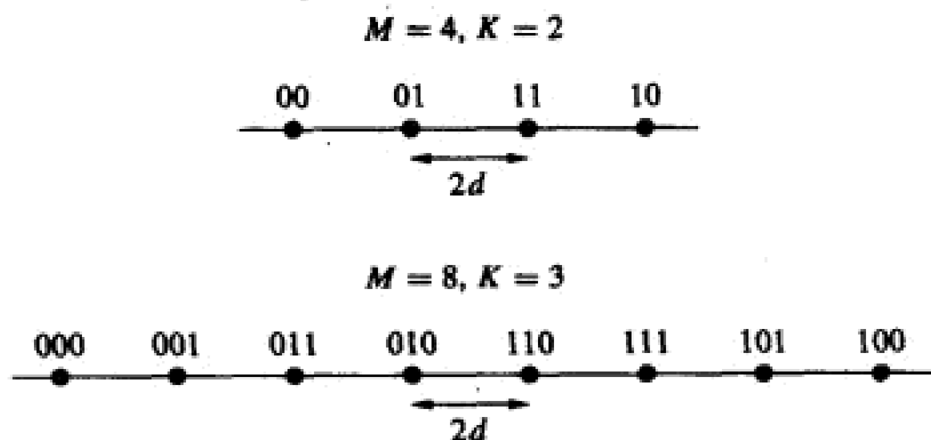


图5-12　MPAM星座的格雷码

例5.4：对于矩形脉冲 $g(t)=\sqrt{2/T_s}, 0\leqslant t<T_s$ ，求4PAM调制的平均能量。

解：4PAM调制的信号幅度为$A_i=\{-3d, -d, d, 3d\}$，所以平均能量为

145

$$\bar{E}_s=\frac{d^2}{4}(9+1+1+9)=5d^2$$

$M=4$和8时的MPAM的判决域Z_i，$i=1,\cdots,M$以及脉幅$A_i=(2i-1-M)d$如图5-13所示。对任意的M，判决域的数学表示为

$$Z_i=\begin{cases}(-\infty, A_i+d) & i=1\\ [A_i-d, A_i+d) & 2\leqslant i\leqslant M-1\\ [A_i-d,\infty) & i=M\end{cases}$$

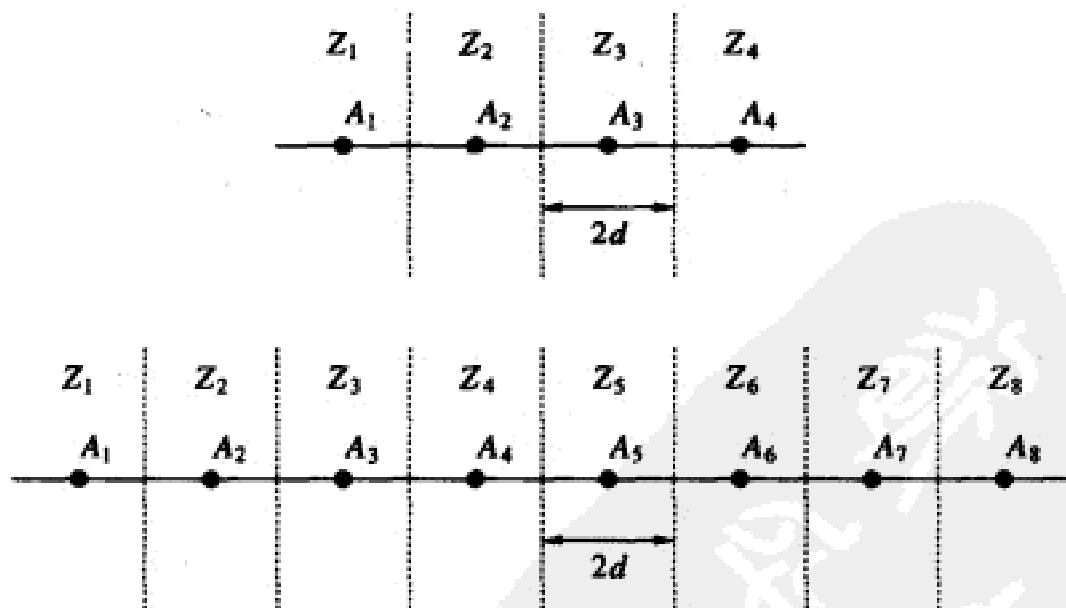


图5-13　MPAM的判决域

1. 原书为Square MQAM，疑有误，此处已改。——译者注

由式（5-52）可知MPAM只有一个基函数$\phi_1(t)=g(t)\cos(2\pi f_c t+\phi)$。所以图5-11所示的相干解调器可以简化为如图5-14所示，其中的多门限设备将r映射到判决域Z_i并输出相应的比特序列$\hat{m}=m_i=\{b_1,\cdots,b_k\}$。

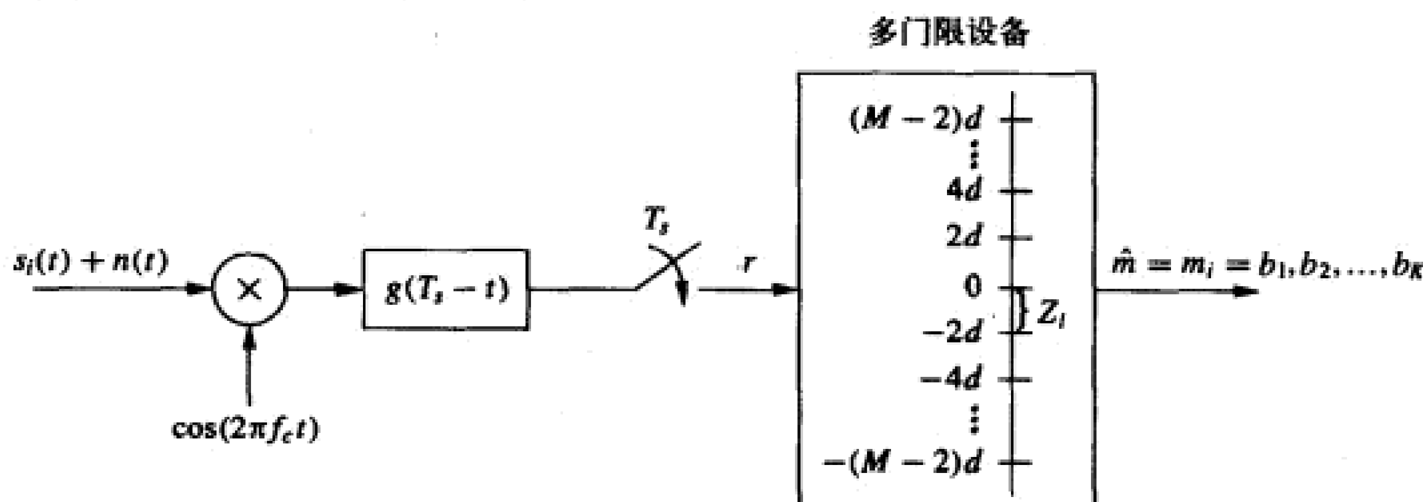


图5-14 MPAM相干解调器

5.3.2 相移键控(MPSK)

MPSK通过相位携带信息。码元间隔$0\leqslant t<T_s$内的发送信号为

$$\begin{aligned}s_i(t)&=\mathrm{Re}\{Ag(t)\mathrm{e}^{\mathrm{j}2\pi(i-1)/M}\mathrm{e}^{\mathrm{j}2\pi f_c t}\}\\&=Ag(t)\cos\left[2\pi f_c t+\frac{2\pi(i-1)}{M}\right]\\&=Ag(t)\cos\left[\frac{2\pi(i-1)}{M}\right]\cos 2\pi f_c t-Ag(t)\sin\left[\frac{2\pi(i-1)}{M}\right]\sin 2\pi f_c t\end{aligned}\tag{5-55}$$

146

星座点（s_{i1}, s_{i2}）由$s_{i1}=A\cos[2\pi(i-1)/M]$及$s_{i2}=A\sin[2\pi(i-1)/M]$给定，$i=1,\cdots,M$。成形脉冲$g(t)$满足式（5-12）和式（5-13）。$\theta_i=2\pi(i-1)/M$（$i=1,2,\cdots,M=2^K$）是携带信息比特的星座点相位。星座点间的最小距离是$d_{\min}=2A\sin(\pi/M)$，其中A是信号能量的函数。2PSK通常叫做BPSK（binary PSK），4PSK通常叫做QPSK（quadrature phase shift keying），4PSK同时也就是4QAM，5.3.3节将介绍MQAM。

MPSK中所有发送信号有相同的能量：

$$E_{s_i}=\int_0^{T_s}s_i^2(t)\,\mathrm{d}t=A^2\tag{5-56}$$

采用矩形脉冲$g(t)=\sqrt{2/T_s},0\leqslant t\leqslant T_s$时，MPSK有恒定的包络，这一点与MPAM、MQAM不同。但矩形脉冲频谱效率低，而高频谱效率的脉冲则使MPSK的包络不再恒定。如MPAM一样，MPSK也常采用格雷映射，信号相位随比特值的变化而作相应改变，如图5-15所示。此时译错一个符号只有一比特错。

MPSK的判决域如图5-16所示。将接收信号用极坐标表示，$r=r_1+\mathrm{j}r_2=r\mathrm{e}^{\mathrm{j}\theta}\in\mathbb{R}^2$，则对任意给定的$M$，其判决域为

$$Z_i=\{r\mathrm{e}^{\mathrm{j}\theta}:2\pi(i-1.5)/M\leqslant\theta<2\pi(i-0.5)/M\}\tag{5-57}$$

147

由式（5-55）可见MPSK包含同相和正交分量，其相干解调器如图5-11所示。对于BPSK，判决域为$Z_1=(r:r>0)$和$Z_2=(r:r\leqslant 0)$（例5.2）。BPSK只有一个基函数$\phi_1(t)=g(t)\cos(2\pi f_c t)$，每码元只传送1比特信息（比特间隔$T_b=T_s$）。这样，BPSK的相干接收机可简化为如图5-17所示。图中门限开关将r映射到与比特值对应的正半部分或负半部分，图中比特值为1映射到星座点$s_1=A$，比特值为0映射到$s_2=-A$。

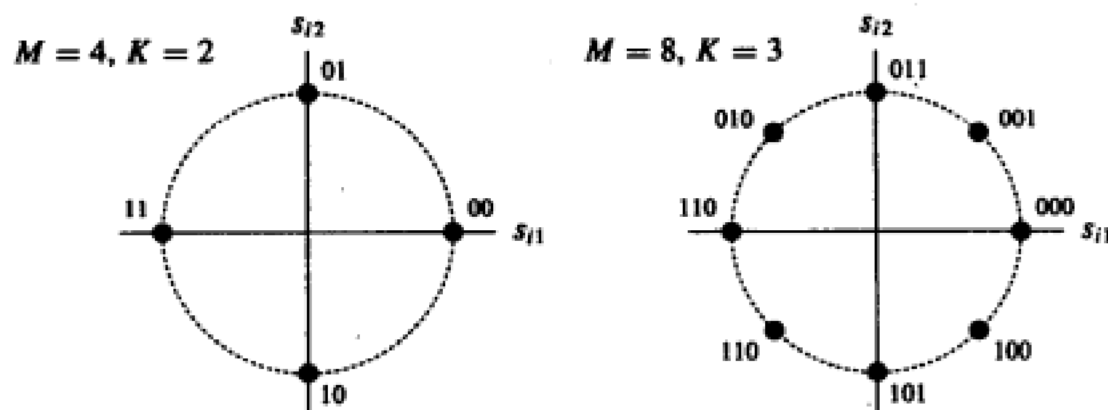


图5-15 MPSK的格雷映射

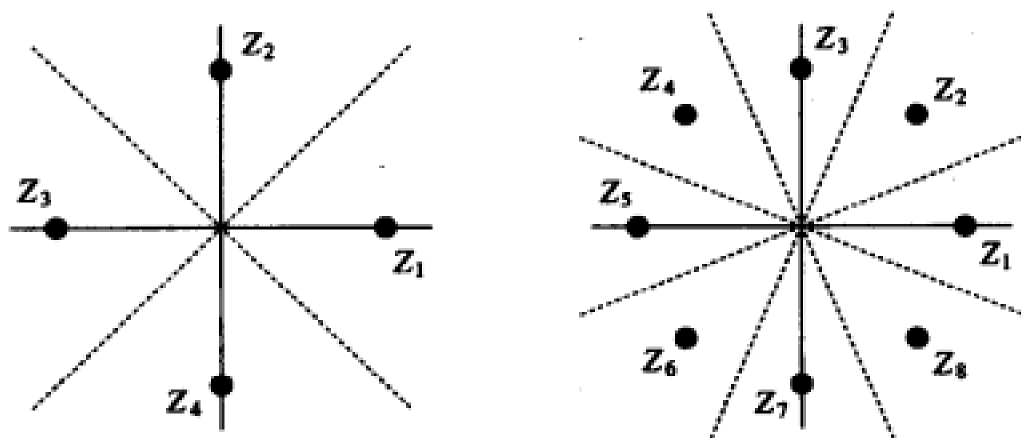


图5-16 MPSK的判决域

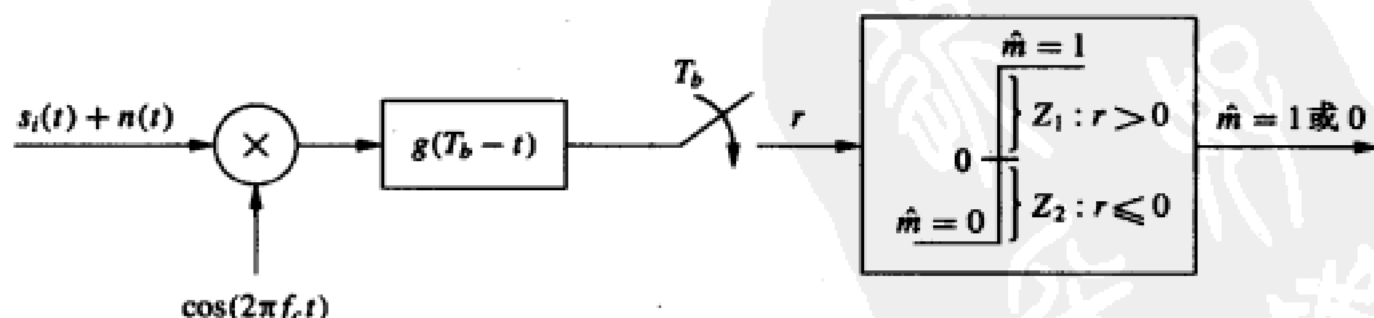


图5-17 BPSK的相干解调

5.3.3 正交幅度调制(MQAM)

MPAM和MPSK都只有一个自由度来携带信息（幅度或相位），MQAM则有两个自由度，其幅度和相位都携带信息。这使得在给定平均能量时，MQAM相比于MPAM和MPSK可以携带更多的信息比特，因此有更高的频谱效率。

MQAM的发送信号为

$$s_i(t) = \mathrm{Re}\{A_i e^{j\theta_i} g(t) e^{j2\pi f_c t}\} \\ = A_i \cos(\theta_i) g(t) \cos(2\pi f_c t) - A_i \sin(\theta_i) g(t) \sin(2\pi f_c t) \quad 0 \leqslant t \leqslant T_s \tag{5-58}$$

其中，脉冲形状$g(t)$满足式（5-12）和式（5-13）。信号$s_i(t)$的能量与MPAM相同，为：

$$E_{s_i} = \int_0^{T_s} s_i^2(t) = A_i^2 \tag{5-59}$$

任意两个星座点的距离为

$$d_{ij} = \|s_i - s_j\| = \sqrt{(s_{i1} - s_{j1})^2 + (s_{i2} - s_{j2})^2} \tag{5-60}$$

148

对正方形星座图，s_{i1}和s_{i2}在$(2i-1-L)d$, $i=1, 2, \cdots, L=2^l$上取值，星座点之间的最小距离与MPAM相同，为$d_{\min}=2d$。大小为L^2的正方形星座的MQAM等价于同相和正交分量都是L进制的MPAM。图5-18示出了常见的方形星座4-QAM和16-QAM。此类星座有$M=L^2=2^{2l}$个星座点，每符号携带$2l$个比特，每维携带l比特，$l=0.5\log_2 M$。对于每维l比特的正方星座的MQAM，平均符号能量P_l与$4^l/3$成正比。每维多传1比特时，$P_{l+1}\approx 4P_l$。也就是说，每维多传1比特，或每符号多传2比特需要增加约6dB的能量来保持星座点间的最小距离不变。

对于QAM信号较难找到很好的星座映射，尤其对于不规则的QAM星座，可能难以建立格雷映射以使相邻的星座点只有1比特不同。图5-19示出了16-QAM的判决域。由式（5-58）可见，MQAM具有同相和正交分量，其相干解调器如图5-11所示。

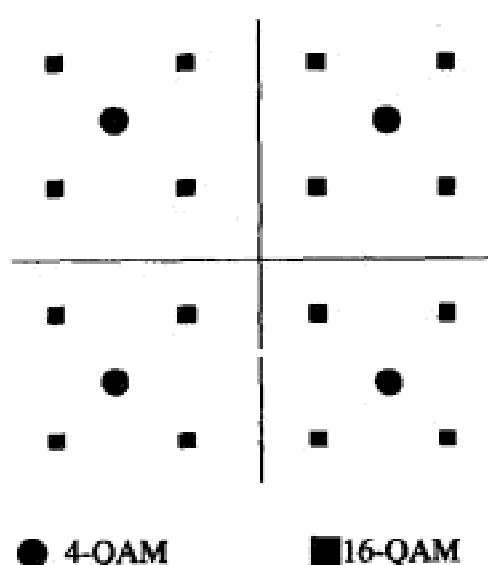


图5-18 4-QAM和16-QAM星座

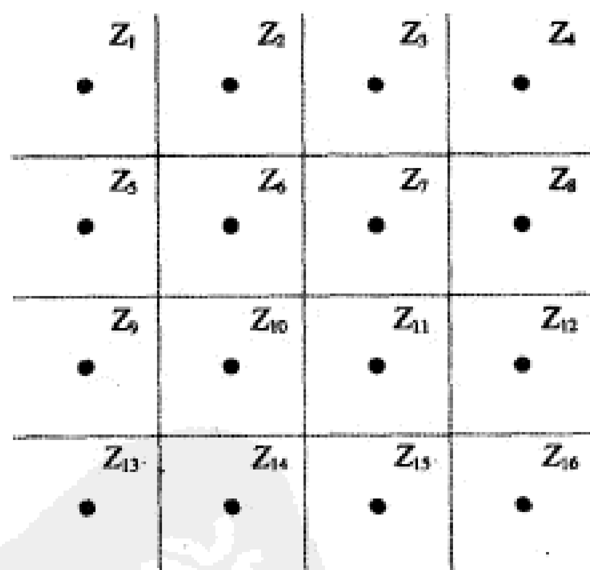


图5-19 MQAM ($M=16$) 的判决域

5.3.4 差分调制

MPSK和MQAM都在相位中携带了信息，所以需要相干解调，即发送载波的相位ϕ_0和接收机的相位ϕ必须要一致。载波相位恢复技术一般会增加接收机复杂性和成本，且易受载波相位漂移的影响。在快变衰落信道中提取相干载波更为困难。5.6节将详细讨论载波恢复问题。载波提取的成本和复杂性促使人们考虑采用其他不要求接收机有相干

149 相位参考的技术，如差分调制。

差分调制是有记忆调制的一种。在有记忆的调制中，码元间隔$[kT_s, (k+1)T_s)$内传输的符号不仅与当前要传送的消息有关，还和以前发送的消息有关。差分调制的基本原理是以前一个符号作为当前符号的参考相位，从而使接收机无需建立相干相位参考。具体的方法就是让前后码元的相位差携带信息。例如对于差分的BPSK（通常记为DPSK），如果在码元间隔$[(k-1)T_s, kT_s)$内传输的符号的相位是$\theta(k-1)=e^{j\theta_i}$，那么在$[kT_s, (k+1)T_s)$间隔内，如果要传0则$\theta(k)=e^{j\theta_i}$，如果要传1则$\theta(k)=e^{j\theta_i+\pi}$。也就是说传0时相位不变，传1时则相位倒$\pi$。类似的，在差分的4PSK调制（DQPSK）中，码元间隔$[kT_s,(k+1)T_s)$内发送符号的相位取决于该时刻要传的信息及前一码元间隔的发送相位。DQPSK的相位变化如表5-1所示。假如$[(k-1)T_s, kT_s)$间隔内发送的相位是$\theta(k-1)=e^{j\theta_i}$，那么在$[kT_s, (k+1)T_s)$间隔内如果要传信息00，则$\theta(k)=e^{j\theta_i}$，也就是把码元间隔$[(k-1)T_s, kT_s]$内的符号重复一遍。如果在码元间隔$[kT_s, (k+1)T_s]$内要传的信息是01，则$\theta(k)=e^{j(\theta_i+\pi/2)}$，如果要传10则$\theta(k)=e^{j(\theta_i-\pi/2)}$，如果传11则$\theta(k)=e^{j(\theta_i+\pi)}$。注意表5-1所示的这种相位变化满足格雷码映射关系，接收端错译为相邻的相位变化时只发生一比特错误。例如，若第k个消息是00，那么k时刻的相位和第$(k-1)$时刻的相位是相同的，设为θ_i。k时刻最容易发生的检测错误是误以为发送相位是$\theta_i\pm\pi/2$，这样解码得到的信息是01或10，与原发送的信息00只差1比特。对于任意M的差分MPSK，都可以对相位变化做格雷编码，因此比特位全为0的信息，其相位不改变；比特位只有一个1、其余为0的信息，其相位改变$2\pi/M$；比特位有两个1、其余为0的信息，其相位改变$4\pi/M$，依此类推。差分调制在MPSK中应用比较普遍，因为差分映射对MPSK比较简单。通过比较复杂的映射，也可以在MQAM中应用差分调制。差分MPSK通常用DMPSK表示，差分BPSK及差分DPSK分别用DPSK及
150 DQPSK表示。

例5.5：假设第$(k-1)$个发送符号为$s(k-1)=Ae^{j\pi}$，第k个符号起的发送信息是101110，求相应的DPSK调制发送的符号。

解：第一个比特1使相位变化π，故$s(k)=A$。下一个比特是0，相位无变化，故$s(k+1)=A$。下一个比特1使相位变化π，$s(k+2)=Ae^{j\pi}$，依此类推。自k时起的发送符号序列是：$A, A, Ae^{j\pi}, A, Ae^{j\pi}, Ae^{j\pi}$。

表5-1　D-QPSK的格雷码映射

比特序列	相位变化	比特序列	相位变化
00	0	10	$-\pi/2$
01	$\pi/2$	11	π

差分调制的解调器如图5-20所示。假设在k时刻发送的信号是$s(k)=Ae^{j(\theta(k)+\phi_0)}$，对应的接收向量为

$$r(k)=r_1(k)+jr_2(k)=Ae^{j(\theta(k)+\phi_0)}+n(k) \tag{5-61}$$

其中$n(k)$是复高斯白噪声。前一时刻的接收向量为

$$r(k-1)=r_1(k-1)+jr_2(k-1)=Ae^{j(\theta(k-1)+\phi_0)}+n(k-1) \tag{5-62}$$

$r(k)$和$r(k-1)$之间的相位差决定发送的是什么。

$$r(k)r^*(k-1)=A^2e^{j(\theta(k)-\theta(k-1))}+Ae^{j(\theta(k)+\phi_0)}n^*(k-1)$$
$$+Ae^{-j(\theta(k-1)+\phi_0)}n(k)+n(k)n^*(k-1) \tag{5-63}$$

如果没有噪声$(n(k)=n(k-1)=0)$，则式（5-63）中只有第一项非零，从中可以求得相位差。图5-20中的相位比较器用来求出这个相位差并输出相应的判决结果。

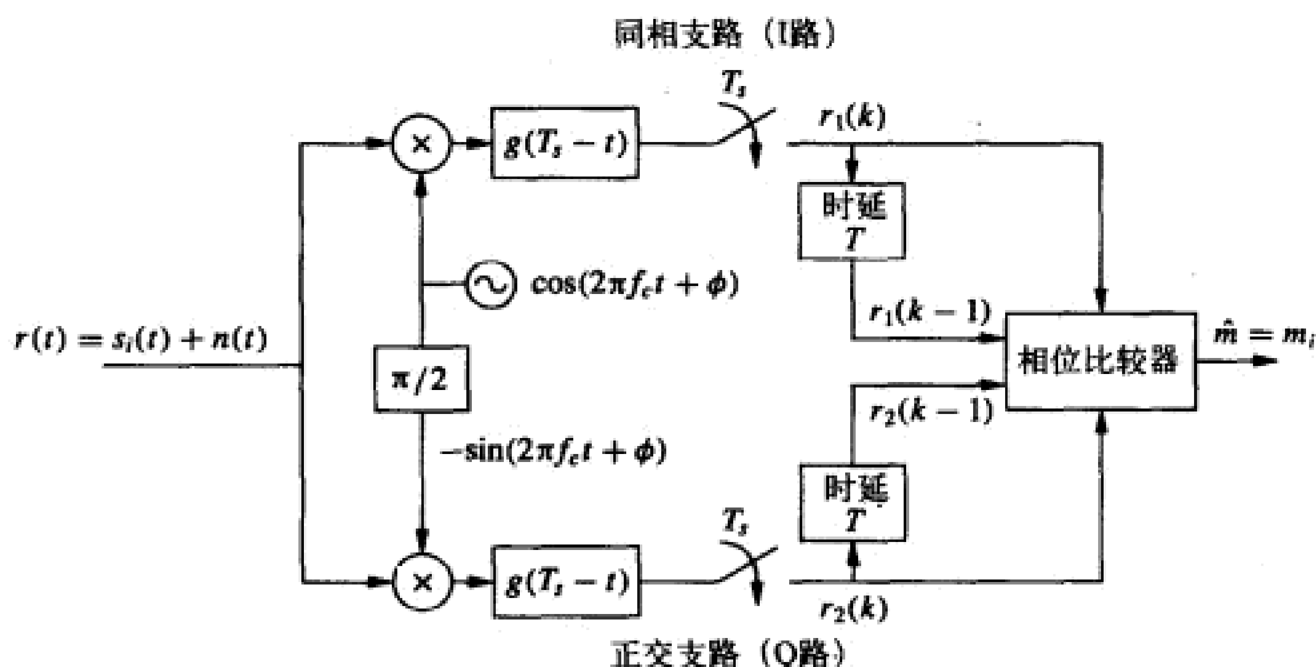


图5-20 差分PSK解调器

差分调制对载波相位的随机漂移并不敏感。但若信道中有显著的多普勒频移，那么相邻码元间隔内的相位可能会不相关，使得前一个符号成为一个很差的相位参考。多普勒频移会使差分调制出现较高的背景误码。详见第6章。 151

5.3.5 星座成形

矩形以及六边形星座的MQAM及MPSK要比正方形和圆形星座的MQAM和MPSK有更好的功率效率。以星座映射的复杂性增加为代价[6]，这些非规则星座可以节约1.3dB功率。最优的星座形状是N维球体，实现这种星座时必须要把它的一个星座点映射为一列2维星座点，这样就可以用图5-10的调制器来实现。文献[6]认为，对于无编码的调制，并不值得使用复杂的球形星座，因为采用纠错编码能以更低的复杂度获得更多的性能增益。不过对于已经使用了某种复杂信道编码的系统，若编码方面即使提高复杂度也无多少潜力可挖的话，可以考虑采用星座成形以得到约1dB的额外增益。文献[6]中深入研究了星座成形，还包括每符号包含非整数比特的星座。

5.3.6 正交偏移

线性调制的符号$s_i=(s_{i1}, s_{i2})$一定落在信号空间四个象限的一个象限中。在时刻kT_s，符号值的变化有可能引起相位发生180度的变化，这将使信号幅度的变化通过零点。经过非线性的放大器或滤波器时，这些相位和幅度的突变会引起失真。将正交分量$g(t)$偏移半个码元时间（如图5-21所示），可以消除或减弱这种突变。这样的方法叫正交偏移，其目的是降低对非线性失真的敏感度。 152

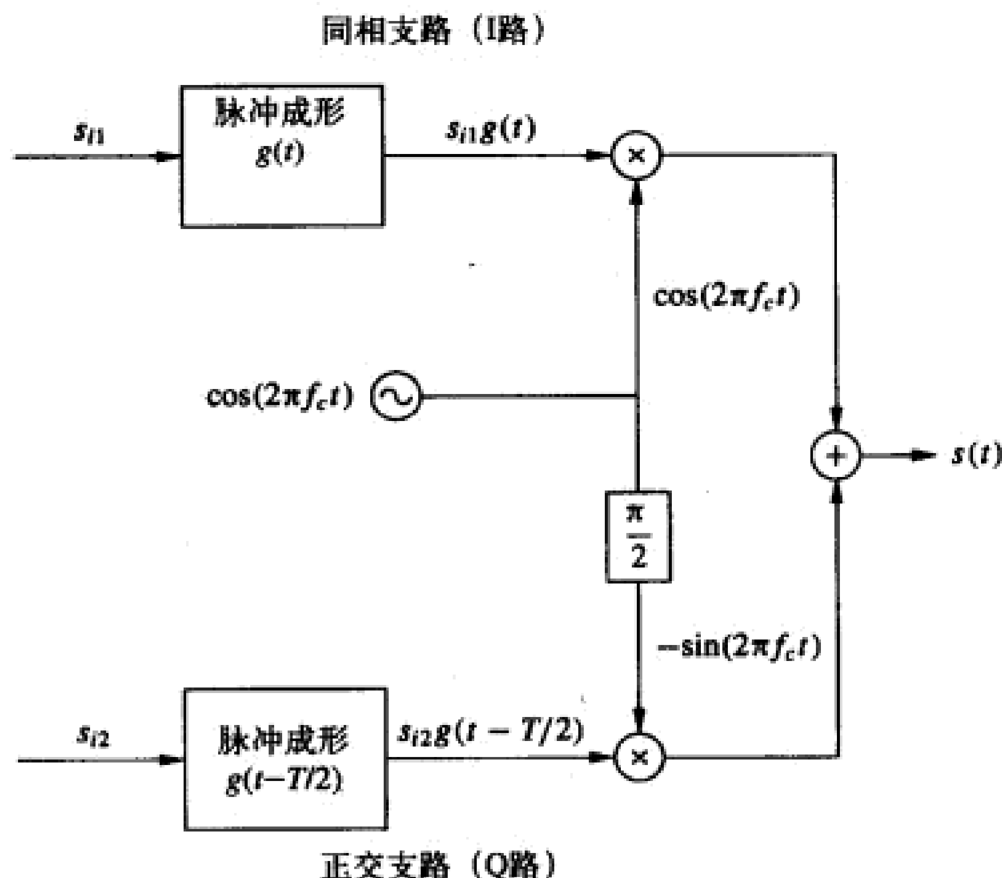


图5-21 带正交偏移的调制器

采用了正交偏移的相位调制通常简写为OMPSK，其中O表示偏移(offset)。例如正交偏移的QPSK称为OQPSK。OQPSK与QPSK在线性放大的情况下具有相同的频谱特征，但在非线性放大时，OQPSK有更高的频谱效率。这是因为在OQPSK信号中，I路和Q路不可能同时变化，所有最大相位变化只是90°。IS-136数字蜂窝系统采用的π/4-QPSK是另一种能避免180°相位变化的方法[7-8]。π/4-QPSK的最大相位变化是135°，相对于OQPSK的90°和QPSK的180°来说，π/4-QPSK在非线性放大时的频谱特性不如OQPSK，π/4-QPSK有一个优点是可以实现差分调制，从而避免提取相干参考相位。差分调制的π/4-QPSK叫π/4-DQPSK，其工作原理是：先将信息比特调制为DQPSK，其星座是QPSK，然后每隔一个码元间隔，将星座图旋转π/4相位。这种旋转所产生的效果和OQPSK中的延时一样的：前后符号变化时不可能产生180°的相位变化。这样减少了码元传输的幅变振荡，增强了信号抗噪声及衰落的能力。

5.4 频率调制

频率调制通过信号的频率携带信息。信号$s(t)$在每个码元时间包含$\log_2 M$个比特消息，对于第i个消息，发送信号为$s_i(t)=A\cos(2\pi f_i t+\phi_i)$，$\phi_i$是第$i$个载波的相位。信号空间表示为$s_i(t)=\sum_j s_{ij}\phi_j(t)=A\phi_i(t)$，其中$s_{ij}=A\delta(i-j)$，$\phi_j(t)=\cos(2\pi f_j t+\phi_j)$。基函数对应不同的频率的载波，在每个码元间隔内发送一个基函数。若$\phi_i=\phi_j$，则能使基函数正交的最小载波频率间隔是$\Delta f=\min_{ij}|f_j-f_i|=1/(2T_s)$，若$\phi_i\neq\phi_j$，这个最小间隔是$1/T_s$。

153

频率调制因为信息被调制到频率中，所以发送信号有恒定的包络A。这样就可以用功率效率高的非线性放大器，并且对信道和硬件引起的幅度失真不太敏感。其代价是频谱效率低：作为一种非线性调制，它占用的带宽要比5.3节介绍的幅度/相位调制更多。

最简单的频率调制器如图5-22所示，相应的解调器如图5-23。这个解调器要求第j个载波的相位必须要和发送的第j个载波同相，这一点类似于幅度/相位调制中的相干载波。5.4.3节将给出一种不需要相干载波的接收机。频率调制的另一个问题是，图5-22中不同的载波有不同的相位，即$\phi_i \neq \phi_j$，$i \neq j$，这样码元变换会造成发送信号的相位不连续。这种不连续会大大增加信号带宽，所以实际采用的是5.4.2节介绍的连续相位频率调制。 [154]

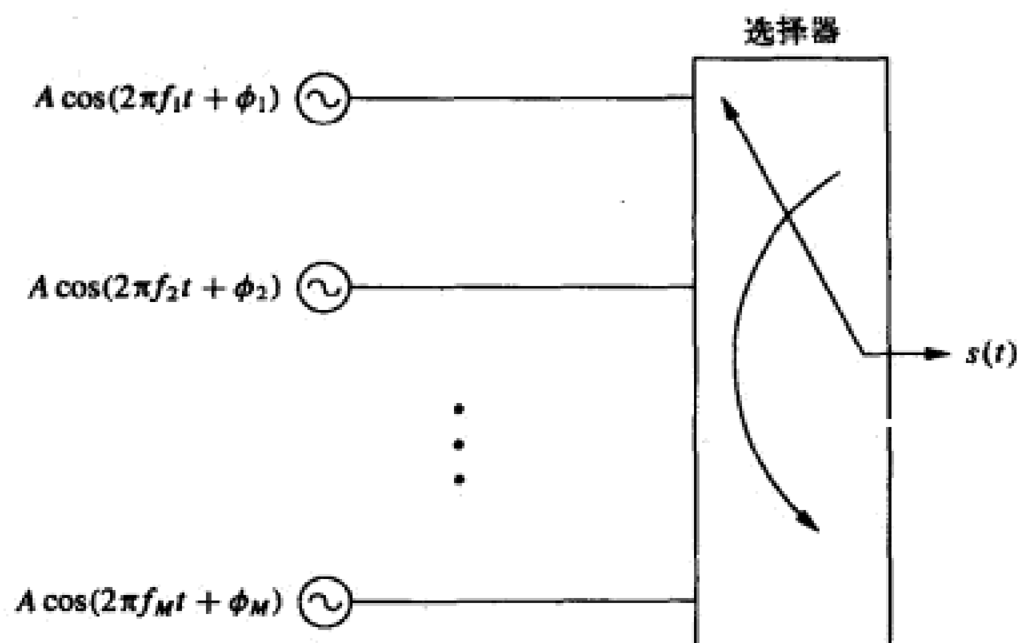


图5-22 频率调制器

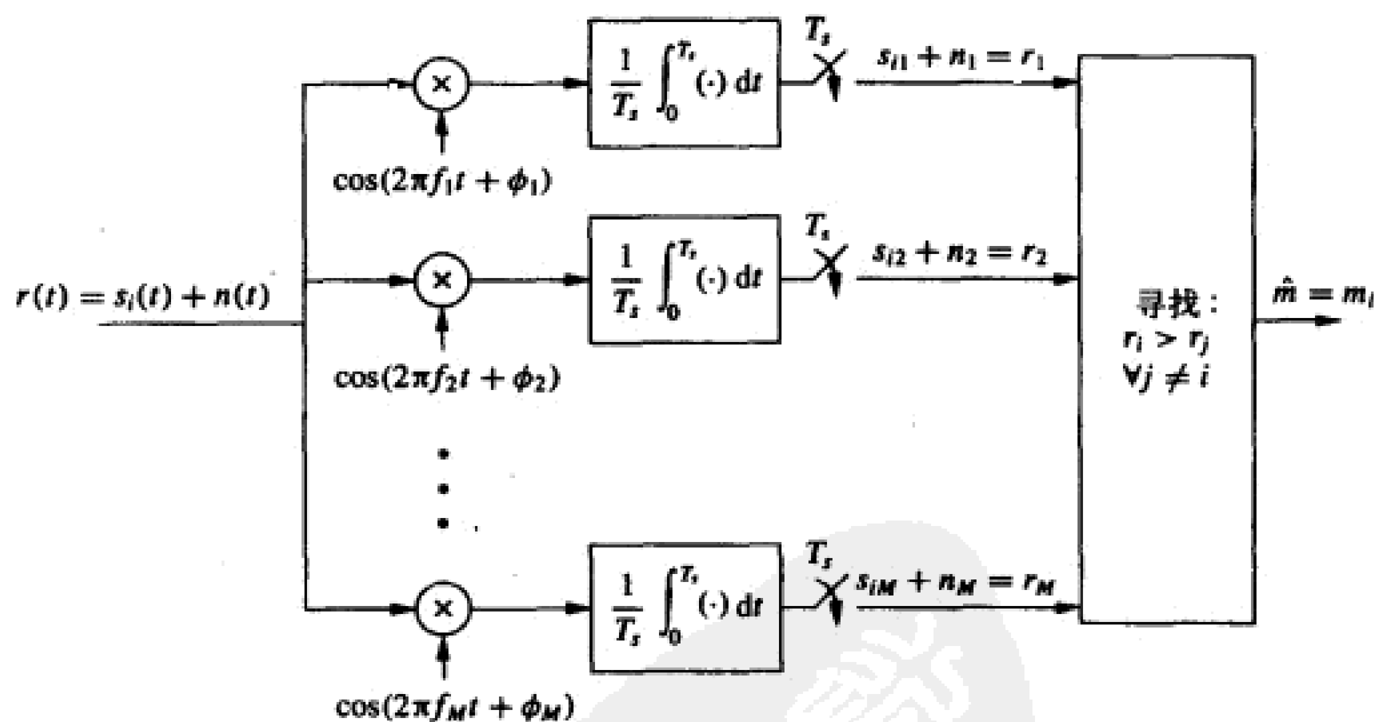


图5-23 频率解调器(相干)

5.4.1 移频键控(FSK)和最小移频键控(MSK)

MFSK的调制信号为

$$s_i(t) = A\cos[2\pi f_c t + 2\pi\alpha_i \Delta f_c t + \phi_i] \quad 0 \leqslant t < T_s \tag{5-64}$$

其中$\alpha_i=(2i-1-M)$, $i=1, 2,\cdots, M=2^K$。载波间的最小频率间隔是$2\Delta f_c$。MFSK由M个基函数 $\phi_i(t)=\sqrt{2/T_s}\cos[2\pi f_c t+2\pi\alpha_i\Delta f_c t+\phi_i]$ $(i=1,\cdots,M)$ 组成，其中$\sqrt{2/T_s}$是归一化因子以使$\int_0^{T_s}\phi_i^2(t)=1$。在一个码元间隔内只能发送一个基函数。

产生MFSK信号的一种简单方法如图5-22所示，M个振荡器工作在不同频率$f_i=f_c+\alpha_i\Delta f_c$上，调制器每隔时间$T_s$就在各振荡器间切换一次。各振荡器的相位偏差将导致调制器输出信号相位的不连续性，这会导致频谱展宽。下一节将介绍一种连续相位的FSK调制器。MFSK的相干检测器如图5-23所示，对于二元调制可简化为图5-24。在图5-24中，判决器在输入大于0时输出为1，在输入小于0时输出为0。

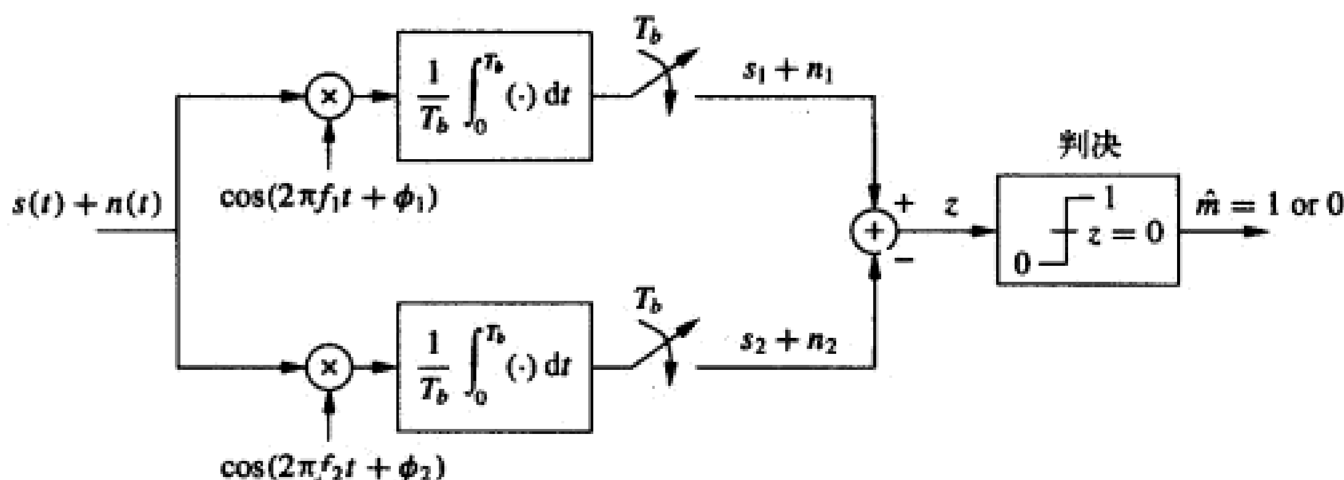


图5-24 2FSK解调器

MSK是二进制FSK的特例，它的两个载波满足$\phi_1=\phi_2$且频率间隔是$2\Delta f_c=1/2T_s$，这是满足正交性的最小间隔。MSK是最小带宽的FSK。

155

5.4.2 连续相位移频键控(CPFSK)

另一种产生MFSK的方法是像模拟调频那样，用数字基带信号去调制单频载波，这种方法可以消除相位不连续，这种情况的已调信号为

$$s_i(t)=A\cos\left[2\pi f_c t+2\pi\beta\int_{-\infty}^{t}u(\tau)\,d\tau\right]=A\cos[2\pi f_c t+\theta(t)] \tag{5-65}$$

其中$u(t)=\sum_k\alpha_k g(t-kT_s)$是5.3.1节所述的MPAM信号。发送信号$s(t)$的相位$\theta(t)$显然是连续的，因此这种MFSK被称为连续相位的FSK（continuous-phase FSK, CPFSK）。

由卡松公式[9]，$s(t)$的带宽近似为

$$B_s\approx 2M\Delta f_c+2B_g \tag{5-66}$$

其中B_g是MPAM信号$u(t)$中成形脉冲$g(t)$的带宽。相比之下，成形脉冲同为$g(t)$的线性调制的带宽为$B_s=2B_g$，可见CPFSK调制占用的频带比线性调制多$M\Delta f_c\geqslant M/(2T_s)$。相对于线性调制，CPFSK在频谱利用率上的损失量随着数据速率的增加而增加，因为数据速率由每符号携带的比特数$K=\log_2M$和符号速率$R_s=1/T_s$决定。

CPFSK的相干解调可以是逐符号的，也可以按符号串进行。由于每个码元间隔内发送的信号和先前发送的有关系，因此同时对所有符号进行检测是最佳的。不过序列检测需要很大的存储器和计算能力。详见文献[10]的5.3节。

5.4.3 FSK的非相干检测

FSK相干解调器需要对每个载波都进行相干载波恢复，这在实现上比较困难而且昂贵。另一种方法是检测每个频率上的信号能量，若第i个频率能量最大，接收机就输出消息m_i。这种改进的接收机如图5-25所示。

令ϕ_i表示在频率f_i上，发送端与接收端振荡器的相位差，则发送频率为f_i时的信号为：

$$\begin{aligned} s(t) &= A\cos(2\pi f_i t + \phi_i) \\ &= A\cos(\phi_i)\cos(2\pi f_i t) - A\sin(\phi_i)\sin(2\pi f_i t) \quad 0 \leqslant t < T_s \end{aligned} \tag{5-67}$$

156

图5-23所示的相干接收机只能检测出上式中的第一项$A\cos(\phi_i)\cos(2\pi f_i t)$，当相差$\phi_i \approx \pm\pi/2$时，检测结果几乎是零。为了克服这个问题，图5-25所示的接收机将接收信号按照每个频率$f_j(j=1,\cdots,M)$分成M个支路。在每个支路上，用对应频率的非相干的同相载波和正交载波与接收信号相乘，在码元时间内积分、采样、求平方。对于第j个支路，记同相分量的平方输出为$A_{jI}+n_{jI}$，正交分量的平方输出为$A_{jQ}+n_{jQ}$，其中n_{jI}和n_{jQ}是噪声分量。如果$i=j$，则$A_{jI}=A^2\cos^2(\phi_i)$、$A_{jQ}=A^2\sin^2(\phi_i)$。如果$i\neq j$，则$A_{jI}=A_{jQ}=0$。不考虑噪声时，判决器的第i路输入为$A^2\sin^2(\phi_i)+A^2\cos^2(\phi_i)=A^2$，与$\phi_i$无关，而判决器所有其他路的输入为零。因此，每个码元时间内，如果判决器的第j路输入最大，则判决发送的频率是f_j。

还有一种类似结构，它的每个支路由匹配滤波器、包络检波器和采样器组成，详见文献[1]的6.8节。另外请注意，图5-25所示的非相干接收机仍然需要准确的码元同步用于采样。码元同步将在5.6节介绍。

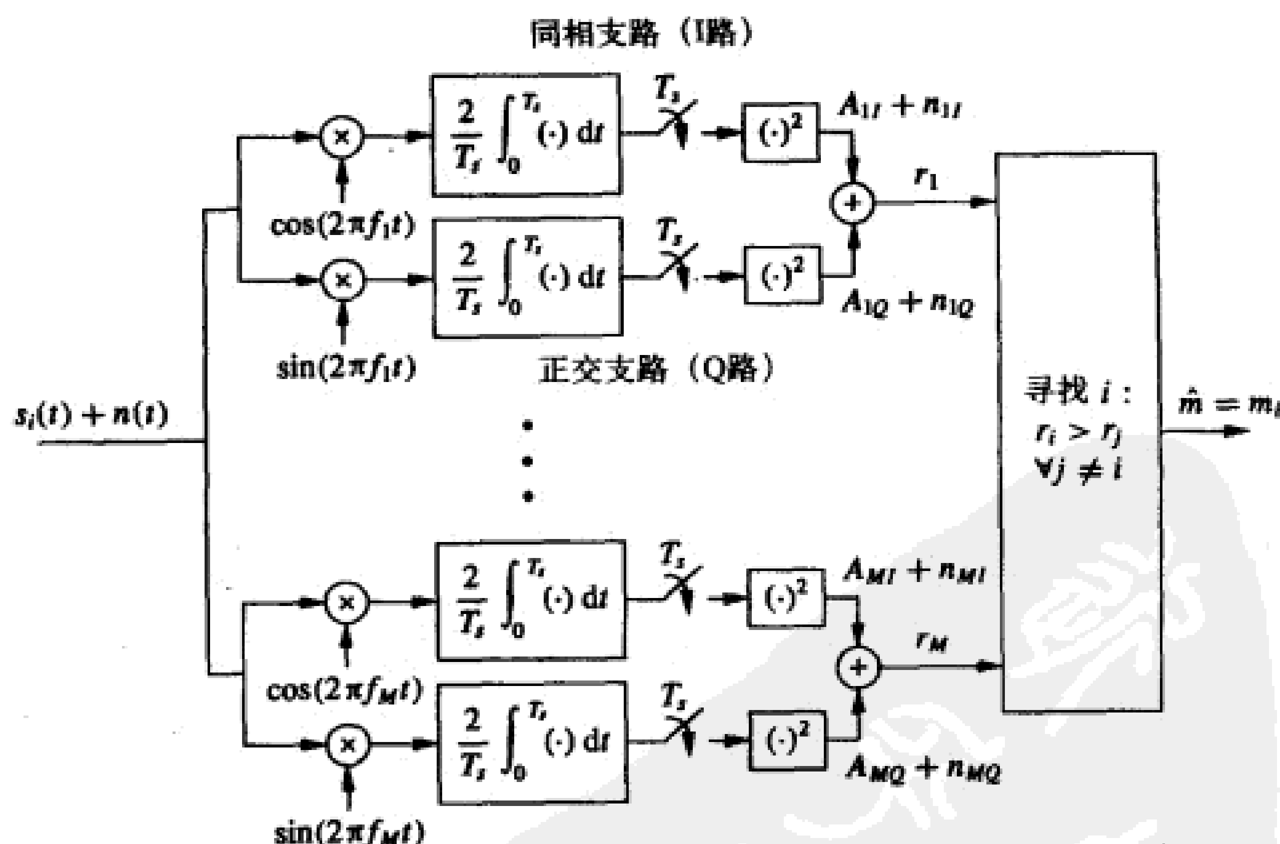


图5-25 非相干FSK解调

5.5 脉冲成形

幅度/相位调制的带宽取决于脉冲$g(t)$的带宽。当$g(t)$是宽度为T_s的矩形脉冲时，信号的包络是恒定的，但矩形脉冲的频谱有较高的旁瓣，会造成对邻近信道的干扰。脉冲成形是一种降低信号旁瓣的方法，不过在进行脉冲成形时必须要注意不能引入码间干扰。注意在接收端采样之前，发送的脉冲$g(t)$经过了与信道冲激响应$c(t)$的卷积及与匹配滤波器$g*(-t)$的卷积，故必须要保证脉冲$p(t)=g(t)*c(t)*g^*(-t)$是无码间干扰的。对于AWGN信道，可设$c(t)=\delta(t)$，则$p(t)=g(t)*g^*(-t)$。对于更一般的$c(t)$的情形，我们将在第11章中讨论。为了没有码间干扰，总体等效脉冲$p(t)$必须满足奈奎斯特准则（Nyquist

157

Criterion），即要求脉冲在过去和未来的采样时刻为零：

$$p(kT_s)=\begin{cases} p_0=p(0) & k=0 \\ 0 & k\neq 0 \end{cases}$$

在频域成为

$$\sum_{l=-\infty}^{\infty} P\left(f+\frac{l}{T_s}\right)=p_0T_s \tag{5-68}$$

下面这些脉冲都满足奈奎斯特准则。

（1）矩形脉冲：$g(t)=\sqrt{2/T_s}$，$0\leqslant t\leqslant T_s$，其总体等效脉冲是三角脉冲

$$p(t)=\begin{cases} 2+2t/T_s & -T_s\leqslant t<0 \\ 2-2t/T_s & 0\leqslant t<T_s \\ 0 & \text{其他} \end{cases}$$

采用矩形脉冲的MPSK调制是恒包络的，但频谱有较高的旁瓣。

（2）余弦脉冲：$p(t)=\sin\pi t/T_s$，$0\leqslant t\leqslant T_s$。余弦脉冲主要用在OQPSK中，其中的Q路有$T_s/2$的时间偏移。它能使旁瓣比矩形脉冲的情形低10dB。

（3）升余弦脉冲：它是为了满足一定的频谱要求而在频域中设计的。脉冲$p(t)$由其傅里叶变换定义为：

$$P(f)=\begin{cases} T_s & 0\leqslant |f|\leqslant \dfrac{1-\beta}{2T_s} \\ \dfrac{T_s}{2}\left[1-\sin\dfrac{\pi T_s}{\beta}\left(f-\dfrac{1}{2T_s}\right)\right] & \dfrac{1-\beta}{2T_s}\leqslant |f|\leqslant \dfrac{1+\beta}{2T_s} \end{cases}$$

其中滚降系数β决定了频谱的衰减速度，如图5-26所示。$P(f)$对应的时域脉冲$p(t)$为

158

$$p(t)=\frac{\sin\pi t/T_s}{\pi t/T_s}\frac{\cos\beta\pi t/T_s}{1-4\beta^2t^2/T_s^2}$$

升余弦脉冲的频域和时域特性如图5-26、图5-27所示。脉冲拖尾以$1/t^3$衰减，因此采样时间误差不会引入严重的码间干扰。升余弦脉冲的一种变化是根升余弦脉冲，其频谱特性是升余弦频谱特性的平方根。根升余弦脉冲的频谱特性比升余弦好，但同步的衰减较慢。根升余弦较慢的时域衰减将会加剧同步误差带来的性能恶化。特别是，取样误差将造成一系列码间干扰分量，其和可能是发散的。

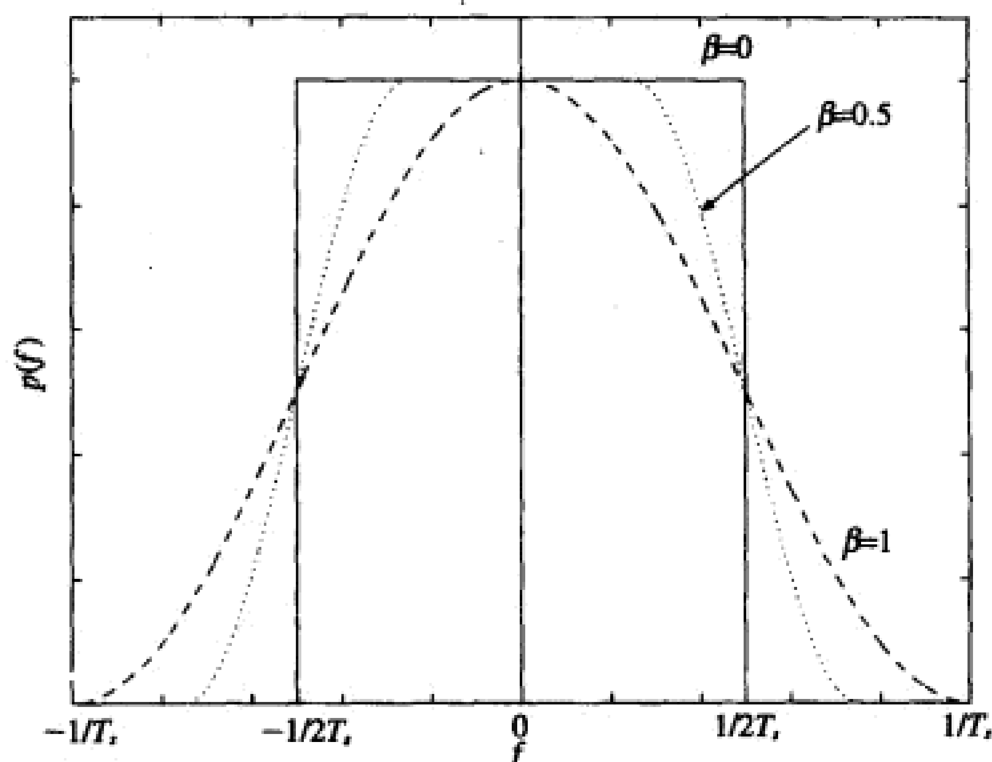


图5-26 升余弦脉冲的频域特性

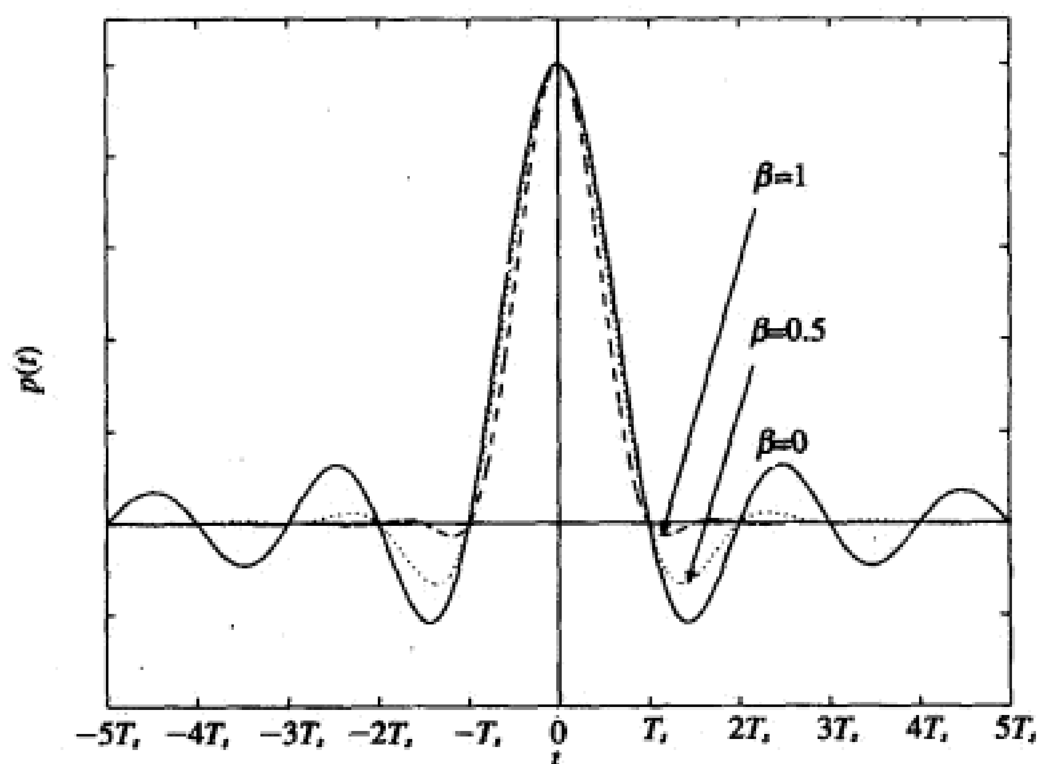


图5-27 升余弦脉冲的时域特性

CPFSK也用脉冲成形来改善频谱效率。CPFSK的做法是对一个MPAM信号进行脉冲成形，再用它来形成FSK信号。CPFSK中脉冲成形常用高斯脉冲

$$g(t)=\frac{\sqrt{\pi}}{\alpha}\mathrm{e}^{-\pi^2t^2/\alpha^2} \tag{5-69}$$

159

其中的α是决定频谱效率的一个参数。CPFSK的频谱取决于$g(t)$的频域特性

$$G(f)=\mathrm{e}^{-\alpha^2f^2} \tag{5-70}$$

α与$g(t)$的3dB带宽有关，关系是

$$\alpha = \frac{0.5887}{B_g} \tag{5-71}$$

显然，增加α将提高频谱效率。

把高斯成形滤波器应用于MSK就得到GMSK。GMSK信号是恒包络，所以有很高的功率效率。同时，高斯脉冲成形对于较大α有较好的频谱特性，所以具有较高的频谱利用率。正是由于这些原因，GSM蜂窝系统采用了GMSK调制。GMSK对话音传输很理想，但对数据业务它不是最好的。高斯脉冲不满足奈奎斯特准则，因此会引入码间干扰，α越大码间干扰也越大。因此通过增加α来改善频谱效率会引入比较大的码间干扰，这种自干扰会导致较高的背景误码。话音业务允许的BER较高，如$P_b \approx 10^{-3}$，一定程度的码间干扰不会影响到这个BER目标。实际中一般认为，若设计$B_gT_s = 0.5$，则所产生的码间干扰对于话音业务是可以接受的。但数据业务要求很低的BER，这就需要对α的最大值及相应的B_g最小值有所限制，也就降低了GMSK的频谱效率。当然可以考虑用均衡来减少码间干扰从而在不影响频谱利用率的前提下达到可接受的BER。不过对于高频谱效率的数据传输，比较普遍的做法是采用线性调制。比如增强的GMS数据业务就采用了线性调制。

5.6　符号同步与载波恢复

符号同步及载波恢复是解调器中一个比较困难的工作。符号同步获取定时信息，用来使接收端对齐采样时间，用于驱动图5-11、图5-23等中的采样电路。载波恢复得到的相位信息用于5.3节、5.4节等中介绍的相干解调。

160

本节简要介绍AWGN信道下实现符号同步与载波恢复的一些标准方法。更广泛来说，定时估计及载波相位估计都属于噪声环境中的信号参数估计。信号估计理论是研究这一问题、建立相关估计方法的理论基础。大多数无线信道除了有加性白高斯噪声外，还要受到时变多径传播的影响。符号同步与载波相位恢复在这种情况下因为很难估计解调前的信号参数将变得十分困难，而且相关的理论基础也比较薄弱。在无线通信系统的性能分析中，一般都是先假设接收机同步于具有平均时延的多径分量[1]，然后按照AWGN信道来处理同步及载波恢复。但在实际中，接收机一般会同步于最强的径或第一个超过一定功率门限的径，其他径会影响接收机的时钟提取及相干载波提取。这种情况在宽带系统，如UWB中更为常见。多载波与扩频系统中在同步及载波恢复方面又会有更多的考虑，这些将在第12章、第13章中介绍。

符号同步及载波恢复的重要性是毋庸置疑的，没有它们，无线通信系统就不能工作。随着传输速率的提高、通信系统自由度的增加（如增加多个天线），同步及载波恢复问题变得更为棘手。虽然这项技术已经研究了许多年，但它仍然在继续发展以适应更高的数据传输、新的系统要求、及更复杂的信道环境。本节只是对码元同步及载波恢复技术作简要介绍。文献[11~12]全面研究了这个问题及相关的性能分析，更详细的讨论还可参考文献[10]的第6章及文献[13]。

1. 这也是为什么时延扩展一般用相对于均值的均方根值表示，见第2章。

5.6.1 具有码元同步及载波恢复的接收机结构

幅度/相位调制系统中的码元同步及载波恢复电路如图5-28所示。对于BPSK，只需保留图中的I路。对于图5-23所示的MFSK相干解调器，每个支路都需要一个载波恢复电路，这样的高复杂性促成了5.4.3节介绍的非相干解调技术。图5-28所示的电路直接用解调前的输入信号进行码元同步及载波恢复。

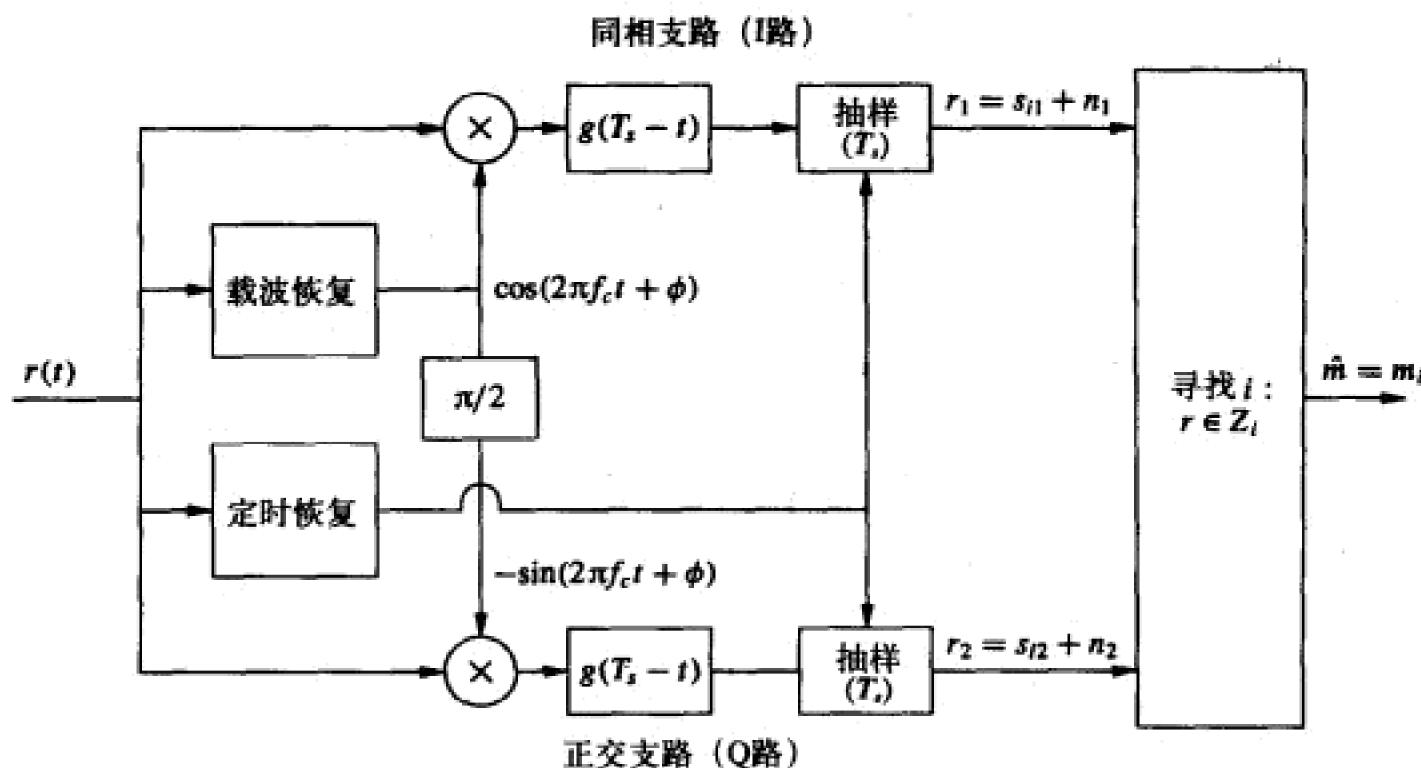


图5-28 具有码元同步及载波恢复的接收机结构

在AWGN信道下，接收信号$r(t)$是发送信号$s(t)$的延迟叠加了加性白高斯噪声$n(t)$：$r(t)=s(t-\tau)+n(t)$，其中τ为随机的传播时延。采用等效基带形式，我们有$s(t)=\mathrm{Re}\{x(t)e^{j\phi_0}e^{j2\pi f_c t}\}$，因此

$$r(t)=\mathrm{Re}\{x(t-\tau)e^{j\phi}e^{j2\pi f_c t}\}+n(t) \tag{5-72}$$

161

其中$\phi=\phi_0-2\pi f_c\tau$是由发送载波相位和传播时延构成的相位。定时恢复就是要估计τ，载波恢复就是要估计ϕ。将这两个未知参数写成向量$\psi=(\phi,\tau)$，则接收信号可写成

$$r(t)=s(t;\psi)+n(t) \tag{5-73}$$

参数估计在有限的时间间隔$T_0\geqslant T_s$内进行，称T_0为观察时间。实际中先是在这个时间内完成一个初始估计，之后是用跟踪环路不断更新这个初始估计。我们将只讨论T_0时间内的初始估计问题，关于跟踪问题请参考文献[11~12]。

对于噪声环境下的信号参数估计，常用方法有两种：最大似然准则(ML)估计和最大后验准则(MAP)估计。ML准则是寻找使$p(r(t)|\psi)$在观察时间T_0内最大的估计值$\hat{\psi}$，MAP准则假设ψ具有某种分布$p(\psi)$，然后寻找使下式在观察时间T_0内最大的估计值$\hat{\psi}$

$$p(\psi\,|\,r(t))=\frac{p(r(t)\,|\,\psi)p(\psi)}{p(r(t))}$$

我们假设没有任何关于估计值$\hat{\psi}$的先验知识，则$p(\psi)$服从均匀分布，此时ML准则与

162 MAP准则等价。

为了表述$p(r(t)|\psi)$，$0\leqslant t<T_0$，我们将观察时间内的$r(t)$用一组正则基函数$\{\phi_k(t)\}$展开为

$$r(t)=\sum_{k=1}^{K} r_k\phi_k(t) \quad 0\leqslant t<T_0$$

由于$n(t)$是白噪声，其均值为零，功率谱密度为$N_0/2$，于是向量$\boldsymbol{r}=(r_1,\cdots,r_k)$在给定$\psi$条件下的概率密度函数为

$$p(\boldsymbol{r}\mid\psi)=\left(\frac{1}{\sqrt{\pi N_0}}\right)^{K}\exp\left[-\sum_{k=1}^{K}\frac{(r_k-s_k(\psi))^2}{N_0}\right] \tag{5-74}$$

其中

$$r_k=\int_{T_0} r(t)\phi_k(t)\,\mathrm{d}t$$

$$s_k(\psi)=\int_{T_0} s(t;\psi)\phi_k(t)\,\mathrm{d}t$$

由此可得

$$\sum_{k=1}^{K}[r_k-s_k(\psi)]^2=\int_{T_0}[r(t)-s(t;\psi)]^2\,\mathrm{d}t \tag{5-75}$$

将其代入式（5-74）可知，使$p(\boldsymbol{r}\mid\psi)$最大等价于使下面的最大似然函数最大

$$\Lambda(\psi)=\exp\left[-\frac{1}{N_0}\int_{T_0}[r(t)-s(t;\psi)]^2\right]\mathrm{d}t \tag{5-76}$$

令此似然函数最大，就可得到载波相位与定时信息的联合ML估计。也可以对载波相位和定时分别进行ML估计，后面几节我们将详细讨论这种分别估计的情形。联合估计比较复杂，请参阅文献[10]的6.4节及文献[11]的第8章至第9章。

5.6.2 最大似然相位估计

本节我们在假设时钟已知的条件下推导最大似然相位估计，此时式（5-76）简化为

163

$$\begin{aligned}\Lambda(\phi)&=\exp\left[-\frac{1}{N_0}\int_{T_0}[r(t)-s(t;\phi)]^2\,\mathrm{d}t\right]\\&=\exp\left[-\frac{1}{N_0}\int_{T_0}x^2(t)\,\mathrm{d}t+\frac{2}{N_0}\int_{T_0}r(t)s(t;\phi)\,\mathrm{d}t-\frac{1}{N_0}\int_{T_0}s^2(t;\phi)\,\mathrm{d}t\right]\end{aligned} \tag{5-77}$$

我们来求使（5-77）最大的相位值$\hat{\phi}$。注意式（5-77）的第一项与ϕ无关，第三项是观察时间内$s(t;\phi)$的能量，假设它也与ϕ无关。这样，使式（5-77）最大的$\hat{\phi}$必然也使下式最大

$$\Lambda'(\phi)=\int_{T_0} r(t)s(t;\phi)\,\mathrm{d}t \tag{5-78}$$

当接收信号是未调制的载波叠加噪声时，即$r(t)=A\cos(2\pi f_c t+\phi)+n(t)$时，可以直接解出$\hat{\phi}$。此时，$\hat{\phi}$必使下式达到最大

$$\Lambda'(\phi)=\int_{T_0} r(t)\cos(2\pi f_c t+\phi)\,\mathrm{d}t \tag{5-79}$$

对ϕ求导并令其为零，则$\hat{\phi}$满足

$$\int_{T_0} r(t)\sin(2\pi f_c t+\hat{\phi})\,\mathrm{d}t=0 \tag{5-80}$$

由此可得

$$\hat{\phi}=-\tan^{-1}\left[\frac{\int_{T_0} r(t)\sin(2\pi f_c t)\,\mathrm{d}t}{\int_{T_0} r(t)\cos(2\pi f_c t)\,\mathrm{d}t}\right] \tag{5-81}$$

我们可以设计一个电路来实现式（5-81）的计算，但在实际中，载波恢复一般是用一个能满足式（5-80）的锁相环（phase lock loop，PLL）来实现的，如图5-29所示。不考虑噪声时，图中积分器的输入为$e(t)=r(t)\sin(2\pi f_c t+\phi)$，积分器输出为

$$z(t)=\int_{T_0} r(t)\sin(2\pi f_c t+\hat{\phi})\,\mathrm{d}t$$

正是式（5-80）的左边。如果$z(t)=0$，则$\hat{\phi}$就是对ϕ的最大似然估计。如果$z(t)\neq 0$，压控振荡器将根据$z(t)$的极性上下调整相位$\hat{\phi}$：$z(t)>0$时通过减小$\hat{\phi}$来减小$z(t)$，$z(t)<0$时通过增大$\hat{\phi}$来增大$z(t)$。实际中又把图5-29所示的积分器换成了一个环路滤波器。滤波器的输入是$e(t)=A\cos(2\pi f_c t+\phi)\sin(2\pi f_c t+\hat{\phi})=A/2[\sin(\hat{\phi}-\phi)+\sin(2\pi f_c t+\hat{\phi}+\phi)]$，输出正比于输入的低频成分：$\sin(\hat{\phi}-\phi)\approx(\hat{\phi}-\phi)$。这里假设了$\hat{\phi}\approx\phi$。如果$\hat{\phi}$和$\phi$相差较大，则$\sin(\hat{\phi}-\phi)\approx(\hat{\phi}-\phi)$不一定成立，$z(t)$的极性不一定会引起正确的相位调整，此时的PLL处在非稳态下。锁相环的优点是它一直在连续不断地调整$\hat{\phi}$以使$z(t)$维持在0附近。这使它能够自动纠正因发送振荡器的频率偏移或传播时延的变化而引起的相位慢变化。锁相环其实是一个反馈控制环路，更多关于锁相环及其性能的介绍见文献[10~11]。

164

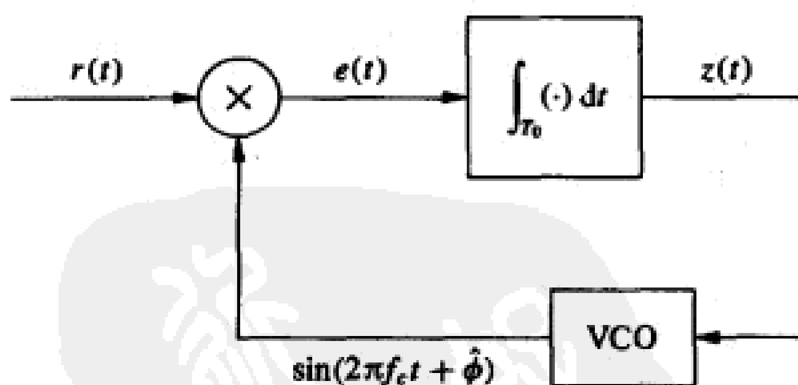


图5-29 用锁相环实现载波恢复(未调制载波)

前面关于锁相环的讨论假设载波未经过调制。在幅度/相位调制中，载波的幅度或相位都要受到数据序列的调制。针对这种情况存在两种方法：一是假设数据序列已知；

二是将其作为随机的，通过统计平均实现相位估计。第一种方法称为非盲估计，达到"数据序列已知"的一种典型方法是发送训练序列。第二种方法称为盲估计，它通过对数据的统计平均使式（5-77）最大似然方程最大化。一种非盲算法是用信息序列的判决结果来去除对载波的调制，再将这样得到的未调载波通过锁相环。这种结构叫判决反馈锁相环，因为它把判决结果反馈给了锁相环。盲算法的载波恢复环路的结构与信息的分布有关。对于较大的星座图，大部分信息分布将涉及被估参数的高度非线性函数，此时若假设信号在每一维都服从高斯分布，可大大简化载波恢复环路。另一种盲算法是先取信号的M次方（对PAM取$M=2$，对MPSK取M为进制数），然后通过中心频率为Mf_c的带通滤波器，再经过锁相环。非线性运算去除了幅度或相位的调制，这样PLL可以提出一个中心频率为Mf_c的未调载波。还有许多其他的盲或非盲算法，其差别在于性能和复杂性的折中。详见文献[10]的6.24节~6.25节及文献[11]。

5.6.3 最大似然定时估计

本节在假设载波相位已知的条件下，推导对时延τ的最大似然估计。由于相位已知，图5-28中的载波恢复及下变频（乘法器）都是理想的，所以只需要考虑$r(t)$和$s(t;\ \tau)$的同相或正交分量即可。记$r(t)$和$s(t;\ \tau)$的同相、正交分量分别为$r_I(t)$、$r_Q(t)$及$s_I(t;\ \tau)$、$s_Q(t;\ \tau)$。因为从I路恢复的时钟也可用于Q路，所以我们只考虑I路。I路的等效低通信号为

165

$$s_I(t;\tau)=\sum_k s_I(k)g(t-kT_s-\tau) \tag{5-82}$$

其中$g(t)$为成形脉冲，$s_I(k)$是第k个码元的同相分量。等效基带接收信号的同相分量为$r_I(t)=s_I(t;\ \tau)+n_I(t)$。如同载波同步一样，定时估计也有两种方法：调制信息已知的非盲估计及调制信息未知的盲估计。

相位ϕ已知时，式（5-76）的似然函数与式（5-77）类似，为

$$\begin{aligned}\Lambda(\tau)&=\exp\left[-\frac{1}{N_0}\int_{T_0}[r_I(t)-s_I(t;\tau)]^2\,\mathrm{d}t\right]\\&=\exp\left[-\frac{1}{N_0}\int_{T_0}r_I^2(t)\,\mathrm{d}t+\frac{2}{N_0}\int_{T_0}r_I(t)s_I(t;\tau)\,\mathrm{d}t-\frac{1}{N_0}\int_{T_0}s_I^2(t;\tau)\,\mathrm{d}t\right]\end{aligned} \tag{5-83}$$

上式中的第一与第三项随τ变化不大，故使式（5-83）最大的$\hat{\tau}$也使下式最大

$$\begin{aligned}\Lambda'(\tau)&=\int_{T_0}r_I(t)s_I(t;\tau)\,\mathrm{d}t\\&=\sum_k s_I(k)\int_{T_0}r(t)g(t-kT_s-\tau)\,\mathrm{d}t=\sum_k s_I(k)z_k(\tau)\end{aligned} \tag{5-84}$$

其中

$$z_k(\tau)=\int_{T_0}r(t)g(t-kT_s-\tau)\,\mathrm{d}t \tag{5-85}$$

对τ求导并令其为零，得$\hat{\tau}$必满足

$$\sum_k s_I(k)\frac{\mathrm{d}}{\mathrm{d}\tau}z_k(\tau)=0 \tag{5-86}$$

对于非盲估计，由式（5-86）可以得到如图5-30所示的估计器。压控时钟（voltage-controlled clock，VCC）的输入为式（5-86）的左边。若此输入为零，则定时估计为$\hat{\tau}=\tau$。若此输入不为零，则将通过调整时钟使VCC的输入为零。这个定时估计电路也是一种反馈控制环路。

166

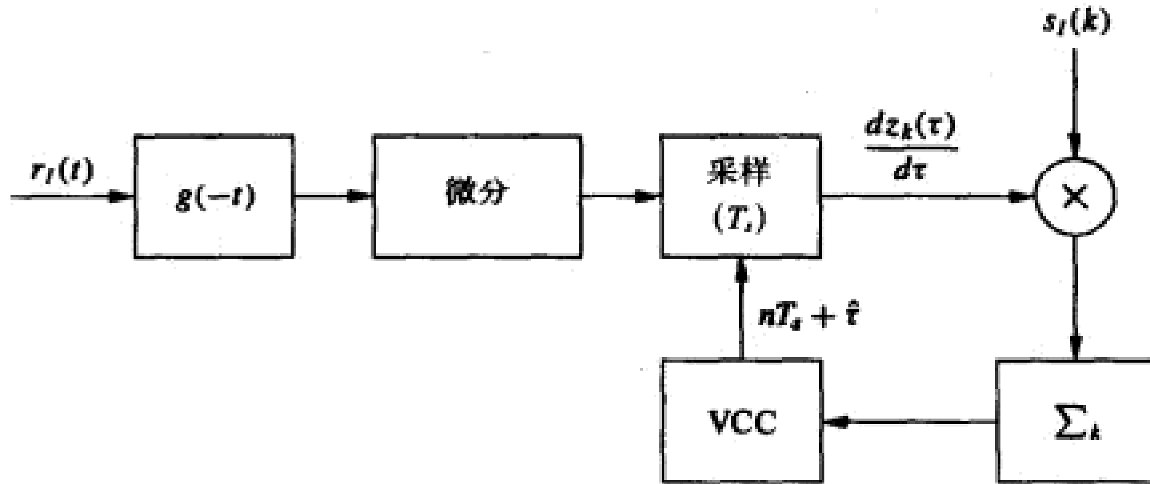


图5-30 非盲定时估计

一种称为超前–滞后门的盲同步器如图5-31所示。它利用了$g(t)$的自相关函数$R_g(\tau)=\int_0^{T_s}g(t)g(t-\tau)dt$的两个性质：$R_g(\tau)=R_g(-\tau)$及$R_g(\tau)$在$\tau=0$时最大。图5-31中上方采样的输入正比于$R_g(\hat{\tau}-\tau+\delta)=\int_0^{T_s}g(t-\tau)g(t-\hat{\tau}-\delta)dt$，下方采样的输入正比于$R_g(\hat{\tau}-\tau-\delta)=\int_0^{T_s}g(t-\tau)g(t-\hat{\tau}+\delta)dt$。若$\hat{\tau}=\tau$，则由于$R_g(\delta)=R_g(-\delta)$，环路滤波器的输入为零，压控时钟将维持当前的正确的定时估计。若$\hat{\tau}>\tau$，则$R_g(\hat{\tau}-\tau+\delta)<R_g(\hat{\tau}-\tau-\delta)$，环路滤波器输入一个负值，使VCC减小$\hat{\tau}$。若$\hat{\tau}<\tau$，则$R_g(\hat{\tau}-\tau+\delta)>R_g(\hat{\tau}-\tau-\delta)$，环路滤波器输入一个正值，使VCC增加$\hat{\tau}$。

更多关于定时估计算法及其性能评估的问题请参考文献[10]的6.2.4节至6.2.5节及文献[11]。

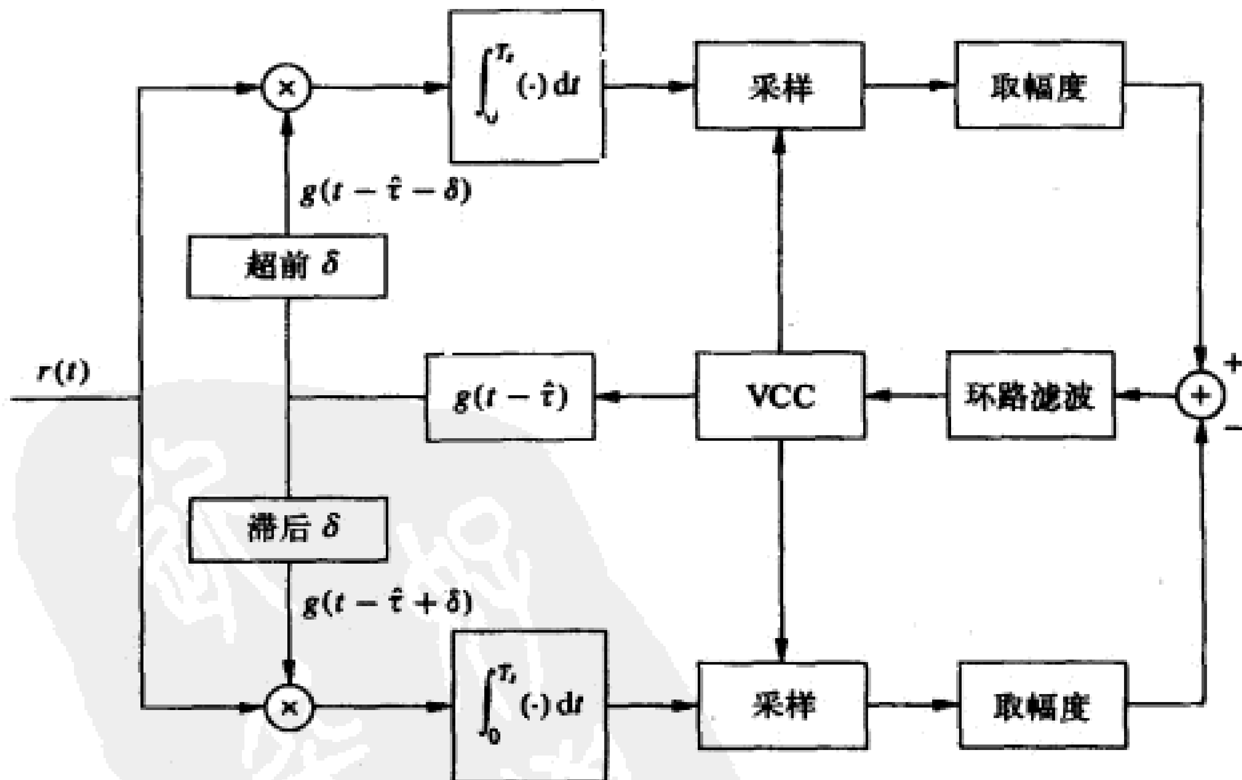


图5-31 超前–滞后门同步器

习题

5.1　利用正则基函数的性质证明：若$s_i(t)$、$s_j(t)$的星座点分别为s_i、s_j，则

$$\|\mathbf{s}_i-\mathbf{s}_j\|^2=\int_0^T(s_i(t)-s_j(t))^2\,\mathrm{d}t$$

167

5.2　对于由$\cos(2\pi t/T)$和$\sin(2\pi t/T)$所张成的信号空间，给出另外一组正交基函数。

5.3　假设有M个互相正交的信号$S_m(t)$，$1\leqslant m\leqslant M$，$0\leqslant t\leqslant T$，其中每个信号的能量为ε。定义一组新的信号为：

$$s'_m(t)=s_m(t)-\frac{1}{M}\sum_{i=1}^{M}s_i(t)\quad 1\leqslant m\leqslant M,\ 0\leqslant t\leqslant T$$

证明这M个新信号的能量都等于$\varepsilon'=(M-1)\varepsilon/M$，并求其中任意两个信号的内积。

5.4　对于图5-32所示的3个信号$\{\phi_1(t),\phi_2(t),\phi_3(t)\}$

(a) 请证明它们是正交的。

(b) 请将$x(t)$展开为$\{\phi_i(t)\}$的线性组合，求出相关系数。其中$x(t)$定义为

$$x(t)=\begin{cases}2 & 0\leqslant t<2\\ 4 & 2\leqslant t\leqslant 4\end{cases}$$

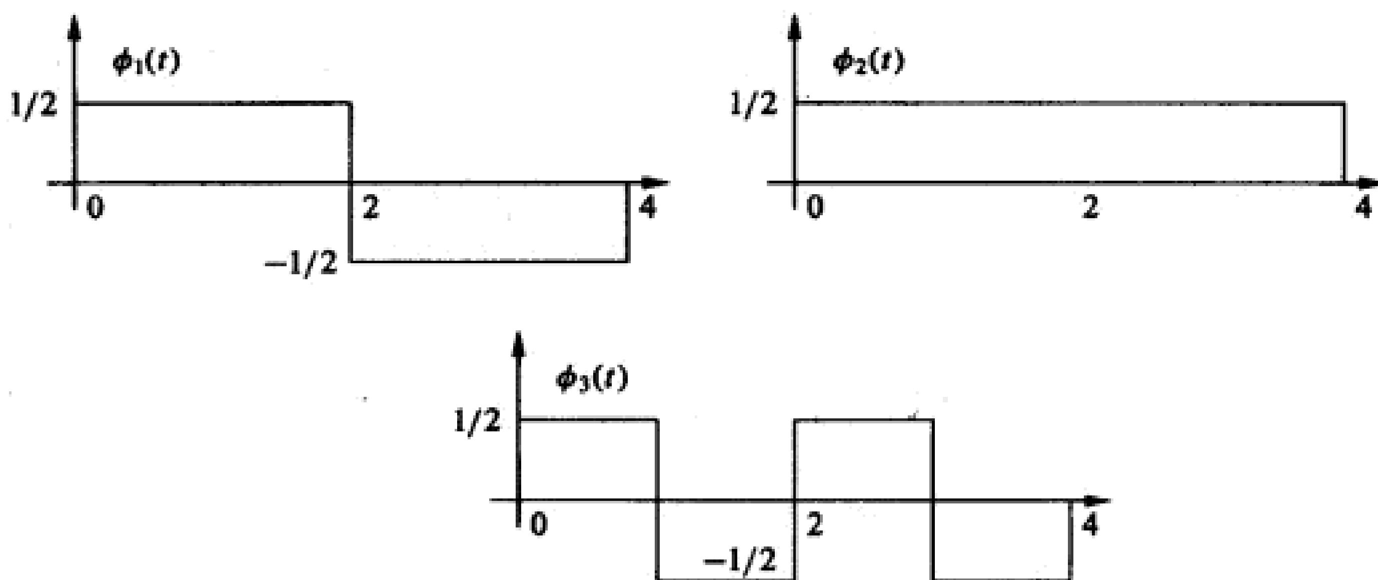


图5-32　题5.4的信号波形

5.5　考虑图5-33所示的4个信号

(a) 求该组信号的维数及相应的基函数。

(b) 用基函数把这几个信号表示为向量。

168

(c) 求各向量间的最小距离。

5.6　假设各信息出现不等概，即$p(m_i)=p_i(i=1,\cdots,M)$，其中p_i不一定等于$1/M$。请导出能使错误率最小的判决域Z_i的数学表达式。并就QPSK星座$s_1=(A_c,0)$、$s_2=(0,A_c)$、$s_3=(-A_c,0)$、$s_4=(0,A_c)$在概率分布为$p(s_1)=p(s_3)=0.2$、$p(s_1)=p(s_3)=0.3$时的情形，给出具体的判决域。

5.7　证明剩余噪声$n_r(t)$与相关器的所有输出r_j独立。注意本题中，r_j在发送s_i条件下是高斯的，$n_r(t)$也是高斯的，高斯随机变量不相关则独立。因此本题也就是要证明$\forall_i$,

$E[n_r(t)r_i]=0$。

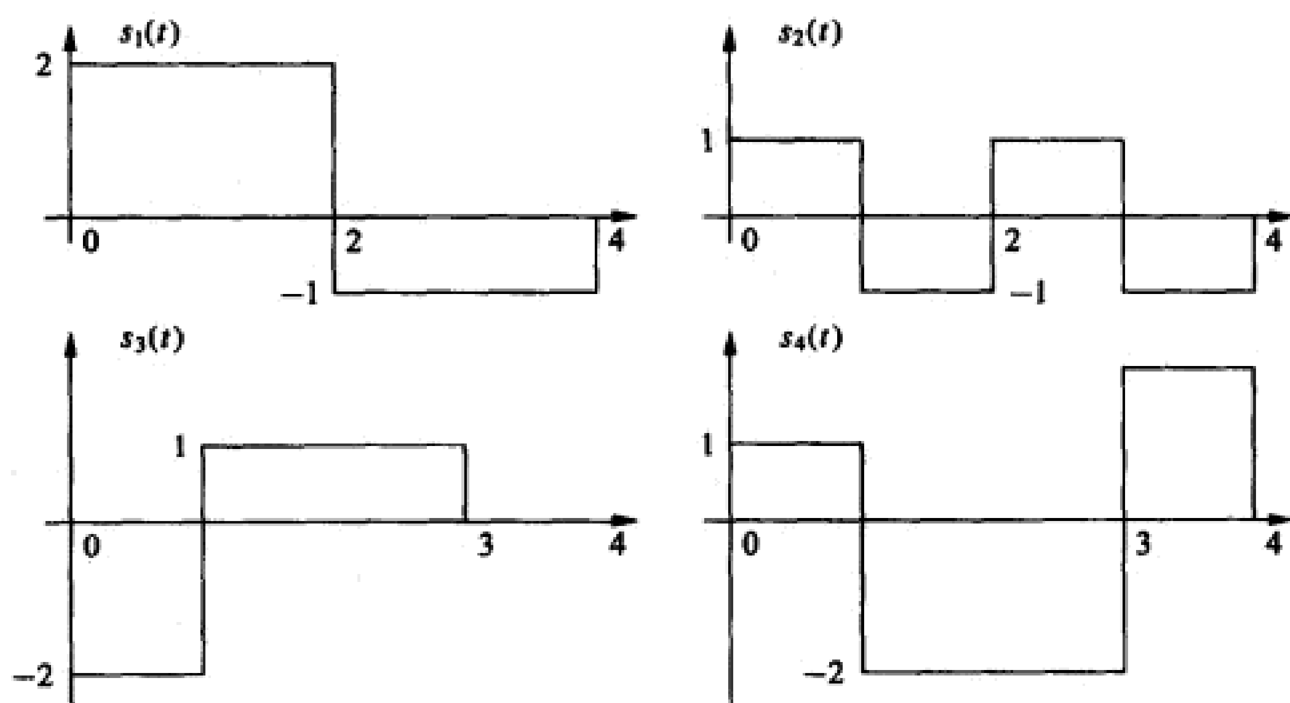


图5-33　题5.5的信号波形

5.8　证明当信号通过一个与之匹配的滤波器时输出信噪比最大。

5.9　求以下信号的匹配滤波器$g(T-t)$, $0\leqslant t\leqslant T$，并求$\int_0^T g(t)g(T-t)\mathrm{d}t$。

(a) 方波脉冲：$g(t)=\sqrt{2/T}$。

(b) sinc脉冲：$g(t)=\mathrm{sinc}(t)$。

(c) 高斯脉冲：$g(t)=\left(\sqrt{\pi}/\alpha\right)\mathrm{e}^{-\pi^2t^2/\alpha^2}$。

5.10　证明图5-4所示的ML接收机等效于图5-7所示的匹配滤波器接收机。

5.11　对于非对称星座$s_1=(A_c,0)$、$s_2=(0,2,A_c)$、$s_3=(-2A_c,0)$、$s_4=(0,-A_c)$，计算式(5-40)、式(5-43)、式(5-44)的界及近似值(5-45)。设$A_c/\sqrt{N_0}=4$。

5.12　如果发送端载波ϕ_0与接收端载波ϕ相位相差$\Delta\phi$，求图5-11中判决器的两个输入。

5.13　设有$d_{\min}=\sqrt{2}$的4-PSK星座。求多增加1比特输出(8-PSK)且仍然保持$d_{\min}$不变（即误码率不变）所需要的能量增量。

5.14　若正方形星座每维有l比特，证明其平均能量P_l与$4^l/3$成正比。若每维增加1个比特，并保持星座点间最小距离不变，证明需要的能量满足关系$P_{l+1}=4P_l$。求$l=2$的P_l并计算具有相同比特/符号及相同最小距离的MPSK及MPAM的平均能量。

5.15　对于差分调制的MPSK，令$\Delta\phi$表示一个码元间隔内信道的相位偏移。在不考虑噪声的情况下$\Delta\phi$需要达到多少才会使接收端的检测发生错误？

5.16　对于差分的8-PSK，列出格雷编码时比特序列和相位变化的对应关系。然后给出比特序列101110100101110对应的调制输出的符号序列，设信息从第k个码元时间开始发送，且第$(k-1)$个码元时间发送的符号为$s(k-1)=A\mathrm{e}^{j\pi/4}$。

5.17　考虑图5-34所示的八进制星座图。

(a) 若8QAM中各星座点间的最小距离为A，求内圆与外圆的半径a、b。

(b) 若8PSK中相邻星座点的间距为A，求半径r。

（c）求这两种星座图的平均发送功率，并作比较。这两个星座图相对的功率增益是多少？（假设发送端符号等概出现）。

（d）对于这两个星座图，有无可能使相邻星座点表示的三比特中只相差一比特？

（e）如果比特率为90Mbit/s，求符号速率。

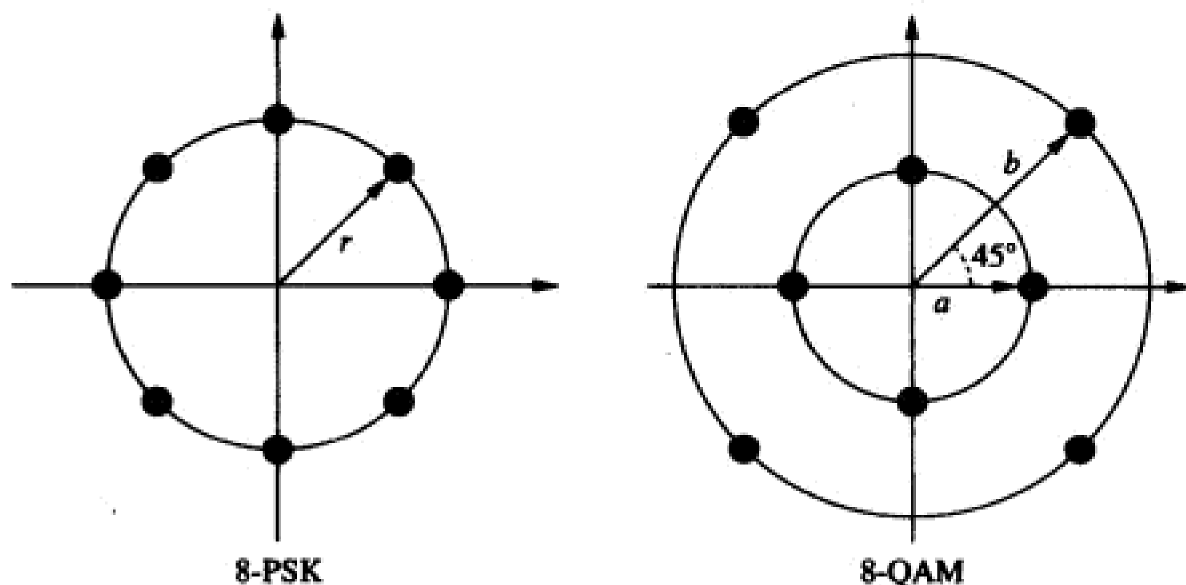


图5-34　题5.17的八进制星座图

5.18　π/4-QPSK调制可看作是两个QPSK系统，它们的星座图相对旋转了π/4。

（a）画出π/4-QPSK的信号空间图。

（b）按格雷码规则标出每个星座点对应的比特序列。

（c）求比特序列0100100111100101通过π/4-QPSK发送的符号序列。

（d）在π/4-DQPSK调制下重做(c)。假设I路所传最后一个符号相位为π，Q路最后一符号相位为−3π/4。

5.19　证明FSK中使载波$\cos(2\pi f_j t)$与$\cos(2\pi f_i t)$正交的最小频率间隔为$\Delta f=\min_{ij}|f_j-f_i|=0.5/T_s$。

5.20　证明：无码间干扰的奈奎斯特准则时域条件$p(kT_s)=P_0\delta(k)$与式（5-69）的频域条件等价。

5.21　证明高斯脉冲不满足奈奎斯特准则。

参考文献

[1] S. Haykin, *Communication Systems*, Wiley, New York, 2002.

[2] J. Proakis and M. Salehi, *Communication Systems Engineering*, Prentice-Hall, Englewood Cliffs, NJ, 2002.

170

[3] J. M. Wozencraft and I. M. Jacobs, *Principles of Communication Engineering*, Wiley, New York, 1965.

[4] M. Fitz, "Further results in the unified analysis of digital communication systems," *IEEE Trans. Commun.*, pp. 521–32, March 1992.

[5] R. Ziemer, "An overview of modulation and coding for wireless communications," *Proc. IEEE Veh. Tech. Conf.*, pp. 26–30, April 1996.

[6] G. D. Forney, Jr., and L.-F. Wei, "Multidimensional constellations – Part I: Introduction, figures of merit, and generalized cross constellations," *IEEE J. Sel. Areas Commun.*, pp. 877–92, August 1989.

[7] T. S. Rappaport, *Wireless Communications – Principles and Practice*, 2nd ed., Prentice-Hall, Englewood Cliffs, NJ, 2001.

[8] G. L. Stuber, *Principles of Mobile Communications*, 2nd ed., Kluwer, Dordrecht, 2001.

[9] S. Haykin, *An Introduction to Analog and Digital Communications*, Wiley, New York, 1989.

[10] J. G. Proakis, *Digital Communications*, 4th ed., McGraw-Hill, New York, 2001.

[11] U. Mengali and A. N. D'Andrea, *Synchronization Techniques for Digital Receivers*, Plenum, New York, 1997.

[12] H. Meyr, M. Moeneclaey, and S. A. Fechtel, *Digital Communication Receivers*, vol. 2, *Synchronization, Channel Estimation, and Signal Processing*, Wiley, New York, 1997.

[13] L. E. Franks, "Carrier and bit synchronization in data communication – A tutorial review," *IEEE Trans. Commun.*, pp. 1107–21, August 1980.

171

第6章 无线信道中数字调制的性能

前一章中我们对数字调制进行了讨论，本章讨论数字调制在AWGN信道和平衰落信道中的性能。衡量性能的指标主要有两个：一个是错误率（误码率或误比特率），另一个是中断率（瞬时信噪比低于给定门限值的概率）。无线信道中的平衰落会大大增加平均误比特率或中断率。无线信道出了存在平衰落外，还存在频率选择性衰落和多普勒频移。频率选择性衰落会引起码间干扰（Intersymbol Interference, ISI），码间干扰会造成接收信号的**背景误码**（error floor）。多普勒频移会引起频谱扩展，频谱扩展则会引起邻信道干扰（当用户移动速度较慢时，这种干扰较小）。多普勒频移还会使相隔一个码元周期后的信道相位去相关，这会使差分相移键控（例如DPSK）产生背景误码。本章将分别讨论噪声、平衰落、频率选择性衰落和多普勒频移等对数字调制性能的影响。

6.1 AWGN信道

本节讨论AWGN信道中各种调制的误码性能。我们首先给出信噪比（signal-to-noise power ratio，SNR）的定义，说明它和比特能量（E_b）及符号能量（E_s）的关系。误码率与这些参数密切相关。本节的分析用到了5.1节中介绍的信号空间的概念。

6.1.1 信噪比、每比特能量及每符号能量

在AWGN信道中，发送信号$s(t)=\mathrm{Re}\{u(t)e^{j2\pi f_c t}\}$受到噪声$n(t)$的干扰。$n(t)$是一个白高斯过程，其均值为零、功率谱密度为$N_0/2$。接收信号$r(t)=s(t)+n(t)$。

我们把接收信号的功率P_r与信号$s(t)$带宽内噪声功率的比值叫做接收信噪比。其中接收到的信号功率P_r与发送功率、路径损耗、阴影衰落及多径衰落都有关系（参见第2章、第3章）。噪声的功率取决于所传信号$s(t)$的带宽和噪声的频谱特性。例如，若发送信号$s(t)$的复包络$u(t)$带宽为B，则$s(t)$的带宽是$2B$。白噪声$n(t)$具有平坦的功率谱密度$N_0/2$，所以在带宽$2B$内的噪声功率为$N=N_0/2\times 2B=N_0B$。于是，接收信噪比为

172

$$\mathrm{SNR}=\frac{P_r}{N_0B}$$

在有干扰的系统中，一般用信干噪比SINR（signal-to-interference-plus-noise）来代替信噪比SNR进行误码率计算。如果干扰近似于高斯噪声，这样做是一种合理的近似。接收的SINR为

$$\mathrm{SINR}=\frac{P_r}{N_0B+P_I}$$

其中P_I为干扰的平均功率。

信噪比可以用比特能量E_b或符号能量E_s来表示

$$\mathrm{SNR}=\frac{P_r}{N_0B}=\frac{E_s}{N_0BT_s}=\frac{E_b}{N_0BT_b} \tag{6-1}$$

其中，T_s为码元间隔，T_b为比特间隔（对于二进制调制，有$T_s=T_b$、$E_s=E_b$）。若成形脉冲满足$T_s=1/B$（如$\beta=1$的升余弦脉冲），则对于多电平信号有$\mathrm{SNR}=E_s/N_0$，对二元信号有$\mathrm{SNR}=E_b/N_0$。对任意脉冲，若$T_s=k/B$，其中k是某个常数，则有$k\times\mathrm{SNR}=E_s/N_0$。

常常将$\gamma_s=E_s/N_0$称为符号信噪比，将$\gamma_b=E_b/N_0$称为比特信噪比。从实际性能的角度出发，我们更关心误比特率P_b和比特信噪比γ_b的关系。在多进制调制（如MPAM和MPSK）中，误比特率不仅和误符号率有关，还同比特到符号的映射有关。一般的做法是先求出误符号率P_s同γ_s的关系，再通过一些精确的或者近似的转换方法得到P_b和γ_b的关系。对于常常使用的格雷码映射，高信噪比时可认为每个符号错误对应一个比特错误，此时我们就有了下面的关系：

$$\gamma_b=\frac{\gamma_s}{\log_2 M} \tag{6-2}$$

$$P_b\approx\frac{P_s}{\log_2 M} \tag{6-3}$$

6.1.2 BPSK和QPSK的错误率

首先考虑理想载波同步时相干解调的BPSK。二进制调制的一个符号对应一个比特，所以误符号率和误比特率相等。假设发0的信号为$s_1(t)=Ag(t)\cos(2\pi f_c t)$，发1的信号为$s_2(t)=-Ag(t)\cos(2\pi f_c t)$，其中$A>0$。从式（5-46）可得误码率为

$$P_b=Q\left(\frac{d_{\min}}{\sqrt{2N_0}}\right) \tag{6-4}$$

173

从5.3.2节中我们知道$d_{\min}=\|s_1-s_0\|=\|A-(-A)\|=2A$，而根据式（5-56），$A$和每比特能量的关系是

$$E_b=\int_0^{T_b}s_1^2(t)\,\mathrm{d}t=\int_0^{T_b}s_2^2(t)\,\mathrm{d}t=\int_0^{T_b}A^2g^2(t)\cos^2(2\pi f_c t)\,\mathrm{d}t=A^2 \tag{6-5}$$

这样的信号星座用比特能量来表示就是$s_0=\sqrt{E_b}$、$s_1=-\sqrt{E_b}$，最小距离就是$d_{\min}=2A=2\sqrt{E_b}$。代入式（6-4）得

$$P_b=Q\left(\frac{2\sqrt{E_b}}{\sqrt{2N_0}}\right)=Q\left(\sqrt{\frac{2E_b}{N_0}}\right)=Q(\sqrt{2\gamma_b}) \tag{6-6}$$

QPSK的I路和Q路都是BPSK调制，理想相干解调时这两个BPSK互相正交。这样，I路及Q路的误比特率都等于BPSK的误码率$P_b=Q(\sqrt{2\gamma_b})$。而误符号率就是任何一路出现一比特错误的概率：

$$P_s=1-[1-Q(\sqrt{2\gamma_b})]^2 \tag{6-7}$$

由于码符号能量被均分到同相支路和正交支路，所以$\gamma_s=2\gamma_b$，因此P_s和γ_s的关系是

$$P_s=1-[1-Q(\sqrt{\gamma_s})]^2 \tag{6-8}$$

根据5.1.5节，QPSK的P_s的联合界为：

$$P_s \leqslant 2Q(A/\sqrt{N_0}) + Q(\sqrt{2}A/\sqrt{N_0}) \tag{6-9}$$

将$\gamma_s=2\gamma_b=A^2/N_0$代入可得

$$P_s \leqslant 2Q(\sqrt{\gamma_s}) + Q(\sqrt{2\gamma_s}) \leqslant 3Q(\sqrt{\gamma_s}) \tag{6-10}$$

闭式界（5-44）变成

$$P_s \leqslant \frac{3}{\sqrt{2\pi\gamma_s}}\exp[-0.5\gamma_s] \tag{6-11}$$

将星座点的最小 $d_{\min}=\sqrt{2A^2}$ 代入式（5-45）可得最近邻近似为

$$P_s \approx 2Q(\sqrt{A^2/N_0}) = 2Q(\sqrt{\gamma_s}) \tag{6-12}$$

此外还可以注意到，每个QPSK码元包含2个比特，因此格雷映射的QPSK近似有$P_b \approx$ 174 $P_s/2$。

例6.1：假设$\gamma_b=7\text{dB}$，求格雷编码的QPSK的误比特率P_b和误符号率P_s。比较P_b的精确值与用$P_b\approx P_s/2$得到的近似值。用$\gamma_s=2\gamma_b$计算最近邻界，并将其与P_s的精确值作比较。

解：$\gamma_b=10^{7/10}=5.012$，因此

$$P_b = Q(\sqrt{2\gamma_b}) = Q(\sqrt{10.024}) = 7.726\times10^{-4}$$

真实误码率为

$$P_s = 1-[1-Q(\sqrt{2\gamma_b})]^2 = 1-[1-Q(\sqrt{10.02})]^2 = 1.545\times10^{-3}$$

按$P_b\approx P_s/2$得到的近似误比特率是7.723×10^{-4}，与真实值相当接近。误符号率的最近邻近似为

$$P_s \approx 2Q(\sqrt{\gamma_s}) = 2Q(\sqrt{10.024}) = 1.545\times10^{-3}$$

和真实值相同。

6.1.3　MPSK的误码率

在MPSK的星座图中，$s_{i1}=A\cos\left[\frac{2\pi(i-1)}{M}\right]$、$s_{i2}=A\sin\left[\frac{2\pi(i-1)}{M}\right]$，$A>0$，$i=1,\cdots,M$。符号能量为$E_s=A^2$，因此$\gamma_s=A^2/N_0$。由式（5-57）可知，发送第$i$个星座点时，若接收向量$r=re^{j\theta}$的角度$\theta\notin[2\pi(i-1.5)/M,\ 2\pi(i-0.5)/M]$，就会发生判决错误。对I路和Q路的噪声$n_1$和$n_2$的联合分布进行函数变换，可以得到$\gamma$和$\theta$的联合分布为（详见文献[1]的5.2.7节）：

$$p(r,\theta) = \frac{r}{\pi N_0}\exp\left[-\frac{1}{N_0}(r^2 - 2\sqrt{E_s}r\cos(\theta) + E_s)\right] \tag{6-13}$$

因为误码率只与θ的分布有关，故对γ积分得

$$p(\theta) = \int_0^\infty p(r,\theta)\,\mathrm{d}r = \frac{1}{2\pi}\mathrm{e}^{-\gamma_s\sin^2(\theta)}\int_0^\infty z\exp\left[-\frac{(z-\sqrt{2\gamma_s}\cos(\theta))^2}{2}\right]\mathrm{d}z \tag{6-14}$$

由于对称性，每个星座点的错误率都是一样的，因此我们按发送星座点$s_1=(A,0)$来求误码率P_s：

$$P_s = 1-\int_{-\pi/M}^{\pi/M} p(\theta)\,\mathrm{d}\theta$$
$$= 1-\int_{-\pi/M}^{\pi/M} \frac{1}{2\pi}\mathrm{e}^{-\gamma_s \sin^2(\theta)}\int_0^{\infty} z\exp\left[-\frac{\left(z-\sqrt{2\gamma_s}\cos(\theta)\right)^2}{2}\right]\mathrm{d}z\mathrm{d}\theta \tag{6-15}$$

当$M>4$时，上式无闭式解，只能通过数值方法得到P_s的准确值。

175

MPSK的星座图中，每个星座点都有两个距离为$d_{\min}=2A\sin(\pi/M)$的近邻。这样，由式（5-45）可得P_s的最近邻近似为

$$P_s \approx 2Q\left(\sqrt{2}A\sin(\pi/M)/\sqrt{N_0}\right)=2Q\left(\sqrt{2\gamma_s}\sin(\pi/M)\right) \tag{6-16}$$

这个结果与准确值之间可能会有较大的误差，但在计算的时候，式（6-16）要比式（6-15）的数值积分简单得多。此外，还可以用利用$p(\theta)$的近似式

$$p(\theta)\approx\sqrt{\gamma_s/\pi}\cos(\theta)\mathrm{e}^{-\gamma_s \sin^2(\theta)} \tag{6-17}$$

将其代入式（6-15）也能得到式（6-16）。

例6.2：比较$\gamma_b=15\text{dB}$时8PSK和16PSK的误比特率。先用式（6-16）计算P_s，再用式（6-3）及式（6-2）近似计算P_b。

解：对于8PSK，$\gamma_s=(\log_2 8)\times 10^{15/10}=94.87$。代入式（6-16）得

$$P_s \approx 2Q\left(\sqrt{189.74}\sin(\pi/8)\right)=1.355\times10^{-7} \tag{6-18a}$$

再由式（6-3）得8PSK的误比特率为$P_b=P_s/3=4.52\times10^{-8}$。

对16PSK，$\gamma_s=(\log_2 16)\times10^{15/10}=126.49$。代入式（6-16）得

$$P_s \approx 2Q\left(\sqrt{252.98}\sin(\pi/16)\right)=1.916\times10^{-3} \tag{6-18b}$$

再由式（6-3）得16PSK的误比特率为$P_b=P_s/4=4.79\times10^{-4}$。在相同$\gamma_b$的情况下，16PSK的误比特率要比8PSK的误比特率要大得多。这是很自然的，因为每个16PSK的星座点包含更多的比特，在每比特能量给定的情况下，16PSK星座点之间的距离更小。

上述MPSK误码率的推导中假设了理想相干解调（接收端准确恢复载波相位）。如果接收端恢复的载波存在相位误差，相位误差就会影响到$p(\theta)$的分布，从而影响到误码率P_s的结果。相位误差的大小与载波相位恢复的方法和信噪比有关。关于相干解调中相位误差的影响，详见文献[1]的附录C、文献[2]的4.3.2节及文献[3~4]。这些文献的结果指出，显著的相位偏差会产生背景误码。另外，相位误差对于多进制PSK会造成I路和Q路的相互干扰，所以多进制的PSK相对于BPSK而言，对相位误差更敏感。在快衰落信道中，多径干涉将使信道相位急剧变化，此时相位误差的问题更为突出。信道快变会引起前后码元间信道相位的变化，因此即使对于差分调制，快变的信道也会造成背景误码[5]。除了相位问题外，定时误差也会使MPSK的性能恶化，详见文献[2]的4.3.3节及文献[6]。

176

6.1.4 MPAM和MQAM的误码率

MPAM的星座点是$A_i=(2i-1-M)d$, $i=1, 2, \cdots, M$。其中每个M~2内的点都有两个距

离为$2d$的近邻。发送这些内点时出错的概率就是噪声在任一方向上超过d的概率：$P_s(s_i)=p(|n|>d)$, $i=2,\cdots,M-1$。两个外点都只有一个距离为$2d$的近邻，因而出错概率就是噪声在一个方向上超过d的概率：$P_s(s_i)=p(n>d)=0.5p(|n|>d)$, $i=1, M$。这样，MPAM的误码率为

$$\begin{aligned}P_s &= \frac{1}{M}\sum_{i=1}^{M} P_s(\mathbf{s}_i) \\ &= \frac{M-2}{M}2Q\left(\sqrt{\frac{2d^2}{N_0}}\right)+\frac{2}{M}Q\left(\sqrt{\frac{2d^2}{N_0}}\right)=\frac{2(M-1)}{M}Q\left(\sqrt{\frac{2d^2}{N_0}}\right)\end{aligned} \tag{6-19}$$

由式（5-54）可得MPAM每符号的平均能量为

$$\bar{E}_s=\frac{1}{M}\sum_{i=1}^{M}A_i^2=\frac{1}{M}\sum_{i=1}^{M}(2i-1-M)^2d^2=\frac{1}{3}(M^2-1)d^2 \tag{6-20}$$

这样可将P_s写成$\bar{E}_s$的形式

$$P_s=\frac{2(M-1)}{M}Q\left(\sqrt{\frac{6\bar{\gamma}_s}{M^2-1}}\right) \tag{6-21}$$

再来考虑大小为$M=L^2$的正方形星座的MQAM。这种MQAM调制可以看成I路和Q路各自是MPAM调制，每路的星座点数是L，每路的能量为MQAM的一半，每路的星座点为$A_i=(2i-1-L)d$，$i=1, 2,\cdots$, L。这样，式(6-21)中代入$L=\sqrt{M}$，取$\bar{\gamma}_s$为MQAM的平均符号能量，可得I路或Q路上各自的误码率为

$$P_{s,\text{branch}}=\frac{2(\sqrt{M}-1)}{\sqrt{M}}Q\left(\sqrt{\frac{3\bar{\gamma}_s}{M-1}}\right) \tag{6-22}$$

因为每路各占一半能量，所以式中$\bar{\gamma}_s$前的系数为3，而不是式(6-21)中的6。两路中有一路出错就等于MQAM的符号出错，于是MQAM的误码率为

$$P_s=1-\left(1-\frac{2(\sqrt{M}-1)}{\sqrt{M}}Q\left(\sqrt{\frac{3\bar{\gamma}_s}{M-1}}\right)\right)^2 \tag{6-23}$$

MQAM的最近邻近似需考虑内点和外点的不同。内点有4个距离为$2d$的近邻，外点有2个或3个距离$2d$的近邻。按保守的4来算，最近邻近似误码率为

177

$$P_s\approx 4Q\left(\sqrt{\frac{3\bar{\gamma}_s}{M-1}}\right) \tag{6-24}$$

对于非矩形星座图的MQAM，误符号率存在如下上限

$$P_s\leqslant 1-\left[1-2Q\left(\sqrt{\frac{3\bar{\gamma}_s}{M-1}}\right)\right]^2\leqslant 4Q\left(\sqrt{\frac{3\bar{\gamma}_s}{M-1}}\right) \tag{6-25}$$

与正方星座时的式（6-24）相同。相应的最近邻近似为

$$P_s\approx M_{d_{\min}}Q\left(\frac{d_{\min}}{\sqrt{2N_0}}\right) \tag{6-26}$$

其中$M_{d\min}$为星座图中最大的近邻个数，$d_{\min}$为星座点间的最小距离。

例6.3：设16QAM的$\gamma_b = 15\text{dB}$（$\gamma_s = \gamma_b \log_2 M$），比较式（6-23）的误码率准确值和式（6-24）的近似值，并同例6.2中相同γ_b下16PSK的误码率作比较。

解：平均码元信噪比[1]为

$$r_s = 4 \times 10^{1.5} = 126.49$$

代入式（6-23）得P_s的准确值为

$$P_s = 1 - \left(1 - \frac{2(4-1)}{4} Q\left(\sqrt{\frac{3 \times 126.49}{15}}\right)\right)^2 = 7.37 \times 10^{-7}$$

代入式（6-24）可得P_s的最近邻近似为

$$P_s \approx 4Q\left(\sqrt{\frac{3 \times 126.49}{15}}\right) = 9.82 \times 10^{-7}$$

它比准确值略大，原因在于式（6-24）是按所有星座点的近邻数都是4得到的。例6.2中计算出的16PSK误码率是$P_s \approx 1.92 \times 10^{-3}$，大约比16QAM的误码率高4个数量级。MQAM同时利用幅度和相位携带信息，而MPSK只用到了相位。因此相同M相同γ_b下，MPSK的误码率比MQAM的误码率大很多。在符号能量或比特能量相同的情况下，MQAM星座点有更好的分布，从而更有效地利用了信号的能量，这是它比MPSK性能好的原因。

MQAM的解调器需要同时对信道的幅度和相位进行估计，以免判决域变形。相位估计误差的影响和之前所讨论的MPSK的情形类似。信道幅度用来对判决域进行伸缩以适应发送的星座图，这种伸缩叫*自动增益控制*（Automatic Gain Control, AGC）。如果信道增益估计有误，那么即使没有噪声也会因为判决域不当而引起误码。接收端一般利用导频符号来估计信道增益，但这种方法估计出的信道增益不是理想的，因此，估计误差造成的误码也是难免的。幅度与相位的估计误差对MQAM性能的影响，请参考文献[7]的10.3节及文献[8]。

178

6.1.5 FSK和CPFSK的误码率

先来考虑采用图5-24所示相干解调的二进制FSK的误码率。因为是相干解调，故可忽略载波中的相位移，于是发送信号可表示为

$$s_i(t) = A\sqrt{2T_b}\cos(2\pi f_i t) \quad i = 1, 2 \tag{6-27}$$

因此，$E_b = A^2$、$\gamma_b = A^2/N_0$。判决器的输入为

$$z = s_1 + n_1 - s_2 - n_2 \tag{6-28}$$

如果$z > 0$，判决器输出1，反之则输出0。假设发送$s_1(t)$，那么

$$z \mid 1 = A + n_1 - n_2 \tag{6-29}$$

当$z = A + n_1 - n_2 \leqslant 0$时就会发生判决错误。相应地，如果发送$s_2(t)$，那么

$$z \mid 0 = n_1 - A - n_2 \tag{6-30}$$

1. 原书为energy（能量），疑为作者笔误，此处已改。——译者注

如果$z=n_1-A-n_2>0$，就会发生判决错误。由于n_1和n_2都是均值为0、方差为$N_0/2$的高斯随机变量，且相互独立。因此其差也是零均值的高斯随机变量，方差为$N_0/2+N_0/2=N_0$。这样，当0和1等概发送时，误比特率为

$$P_b=0.5p(A+n_1-n_2\leqslant 0)+0.5p(n_1-A-n_2>0)=Q\left(A/\sqrt{N_0}\right)=Q\left(\sqrt{\gamma_b}\right) \tag{6-31}$$

$M>2$时相干解调的MFSK的误码率推导非常复杂，且无闭式解，请参考文献[2]中的式(4-92)。文献[9]的8.1节推出了非相干解调的MFSK的误码率为

$$P_s=\sum_{m=1}^{M}(-1)^{m+1}\binom{M-1}{m}\frac{1}{m+1}\exp\left[\frac{-m\gamma_s}{m+1}\right] \tag{6-32}$$

CPFSK的误码率不仅与它是否采用相干解调有关，还与它是逐符号译码还是序列译码有关。发送符号的相关性使得误码率分析必须要考虑多个符号，所以CPFSK的误码率分析非常复杂。文献[1]的5.3节对不同结构的CPFSK误码率均有详细的推导。如同线性调制一样，频率误差和定时误差也会使FSK的性能恶化，有关分析见文献[2]的5.2节及文献[10~11]。

179

6.1.6　相干解调的误码率的近似

前面得到的相干解调的误码率准确式或近似式大都具有下面的形式：

$$P_s(\gamma_s)\approx\alpha_M Q\left(\sqrt{\beta_M\gamma_s}\right) \tag{6-33}$$

其中α_M和β_M取决于近似方法和调制类型。对于最近邻近似，α_M是星座图中距离为最小距离$d_{\min}$的近邻个数，β_M是反映$d_{\min}$和平均符号能量之间的关系的常量。表6-1根据前几节的结论列出了对应PSK、QAM和FSK的错误率公式，注意其中α_M和β_M的取值。

表6-1　相干解调时的近似误码率及误比特率

调　制	$P_s(\gamma_s)$	$P_b(\gamma_b)$
BFSK		$P_b=Q(\sqrt{\gamma_b})$
BPSK		$P_b=Q(\sqrt{2\gamma_b})$
QPSK, 4-QAM	$P_s\approx 2Q(\sqrt{\gamma_s})$	$P_b\approx Q(\sqrt{2\gamma_b})$
MPAM	$P_s\approx\frac{2(M-1)}{M}Q\left(\sqrt{\frac{6\bar{\gamma}_s}{M^2-1}}\right)$	$P_b\approx\frac{2(M-1)}{M\log_2 M}Q\left(\sqrt{\frac{6\bar{\gamma}_b\log_2 M}{M^2-1}}\right)$
MPSK	$P_s\approx 2Q\left(\sqrt{2\gamma_s}\sin\left(\frac{\pi}{M}\right)\right)$	$P_b\approx\frac{2}{\log_2 M}Q\left(\sqrt{2\gamma_b\log_2 M}\sin\left(\frac{\pi}{M}\right)\right)$
矩形MQAM	$P_s\approx 4Q\left(\sqrt{\frac{3\bar{\gamma}_s}{M-1}}\right)$	$P_b\approx\frac{4}{\log_2 M}Q\left(\sqrt{\frac{3\bar{\gamma}_b\log_2 M}{M-1}}\right)$
非矩形MQAM	$P_s\approx 4Q\left(\sqrt{\frac{3\bar{\gamma}_s}{M-1}}\right)$	$P_b\approx\frac{4}{\log_2 M}Q\left(\sqrt{\frac{3\bar{\gamma}_b\log_2 M}{M-1}}\right)$

对性能的规定往往更注重误比特率P_b和比特信噪比γ_b之间的关系。对于格雷码映射及高信噪比条件，可用式（6-3）和式（6-2）将P_s和γ_s化为P_b和γ_b。此时式（6-33）化为：

$$P_b(\gamma_b) = \hat{\alpha}_M Q\left(\sqrt{\hat{\beta}_M \gamma_b}\right) \tag{6-34}$$

其中 $\hat{\alpha}_M = \alpha_M / \log_2 M$ 、 $\hat{\beta}_M = (\log_2 M)\beta_M$。后文中会经常用这种代换来从$P_s$和$\gamma_s$函数关系（6-33）得到$P_b$和$\gamma_b$的关系。

6.1.7 差分调制的误码率

差分调制的误码率取决于图5-20中相位比较器输入的相位差。相位比较器先求出下式的相位，再根据它来判决所发送的符号。

$$\begin{aligned} r(k)r^*(k-1) = {} & A^2 e^{j(\theta(k)-\theta(k-1))} + A e^{j(\theta(k)+\phi_0)} n^*(k-1) \\ & + A e^{-j(\theta(k-1)+\phi_0)} n(k) + n(k)n^*(k-1) \end{aligned} \tag{6-35}$$

180

由于对称性，计算误码率时可假设发送的相位差已给定，不妨设$\theta(k)-\theta(k-1)=0$，于是有

$$\begin{aligned} r(k)r^*(k-1) = {} & A^2 + A e^{j(\theta(k)+\phi_0)} n^*(k-1) \\ & + A e^{-j(\theta(k-1)+\phi_0)} n(k) + n(k)n^*(k-1) \end{aligned} \tag{6-36}$$

定义两个新的随机变量：$\tilde{n}(k) = n(k)e^{-j(\theta(k-1)+\phi_0)}$、$\tilde{n}(k-1) = n(k-1)e^{-j(\theta(k)+\phi_0)}$。它们和$n(k)$、$n(k-1)$相比，具有相同的统计特性。于是

$$r(k)r^*(k-1) = A^2 + A(\tilde{n}^*(k-1) + \tilde{n}(k)) + \tilde{n}(k)\tilde{n}^*(k-1) \tag{6-37}$$

式中第一项对应假设的发送相位差为0，第二项和第三项是噪声。当信噪比较大时，第三项远小于第二项，可以忽略。再将剩余的两项除以A得

$$\tilde{z} = A + \text{Re}\{\tilde{n}^*(k-1) + \tilde{n}(k)\} + j\,\text{Im}\{\tilde{n}^*(k-1) + \tilde{n}(k)\} \tag{6-38}$$

定义 $x = \text{Re}\{\tilde{z}\}$， $y = \text{Im}\{\tilde{z}\}$，则 $\tilde{z}$的相位为

$$\theta_{\tilde{z}} = \tan^{-1} y/x \tag{6-39}$$

在假设发送相位差为零的条件下，当$|\theta_{\tilde{z}}| > \pi/M$时就会发生判决错误。求 $p(|\theta_{\tilde{z}}| > \pi/M)$ 的方法和在相干PSK中的方法是一样的，唯一的区别是式(6-38)中多了一项噪声，因此差分PSK的性能比相干PSK大约恶化3dB。

对于DPSK，分析误码率时只需考虑图5-20的同相支路，所以通常在分析中设$x = \text{Re}\{\tilde{z}\}$。假设发送相位差是零，则当$x = A + \text{Re}\{\tilde{n}^*(k-1) + \tilde{n}(k)\} + \tilde{n}(k)\tilde{n}^*(k-1)/Az<0$时就会发生判决错误。其概率可由以下步骤求出：先找出x的特征函数或矩母函数，然后利用拉普拉斯反变换得到x的分布，最后在判决域$x<0$上积分即可。这是一种相当普遍的方法，广泛适用于AWGN信道和衰落信道中各种调制方式和解调方式。我们后面将会用这种方法来计算衰落信道中，有分集及无分集的情况下线性调制的平均误码率。对于DPSK，x的特征函数可以利用复高斯随机变量的广义二次型得到，详见文献[1]的附录B及文献[12]的附录[B]。最终得到的误比特率为

$$P_b = \frac{1}{2} e^{-\gamma_b} \tag{6-40}$$

对于DQPSK，文献[1]的附录B给出了$\tilde{z}$的特征函数，相应的误比特率为

181

$$P_b = \int_b^\infty x \exp\left[\frac{-(a^2+x^2)}{2}\right] I_0(ax)\,dx - \frac{1}{2}\exp\left[\frac{-(a^2+b^2)}{2}\right] I_0(ab) \tag{6-41}$$

其中 $a \approx 0.765\sqrt{\gamma_b}$， $b \approx 1.85\sqrt{\gamma_b}$，$I_0(x)$是修正的零阶第一类贝塞尔函数。

6.2　Q函数的一种等效表示法

在AWGN信道中，各种相干解调的误码率P_s都可以通过近似式（6-33）表示为Q函数的形式。Q函数定义为标准正态随机变量X超过定值z的概率：

$$\mathrm{Q}(z) = p(X \geqslant z) = \int_z^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,dx \tag{6-42}$$

Q函数无闭式解，上式中的积分区间是无限的，函数的变量z同时出现在积分限上，这使Q函数用起来不是很方便。

Craig于1991年提出了Q函数的一种等效表示法[13]：

$$\mathrm{Q}(z) = \frac{1}{\pi}\int_0^{\pi/2} \exp\left[\frac{-z^2}{2\sin^2\phi}\right] d\phi \quad z > 0 \tag{6-43}$$

这种表示法也可以从Weinstein[14]和Pawula[5]等的工作中推导出。等效表示中的积分区间是个与变量z无关的有限区间，被积函数是关于z的高斯函数。这些特点在推导衰落信道的平均误码率时非常有用。

Craig的初衷是为了简化AWGN信道中误码率的计算。比如，采用等效表示法时，BPSK的误比特率为

$$P_b = \mathrm{Q}\left(\sqrt{2\gamma_b}\right) = \frac{1}{\pi}\int_0^{\pi/2} \exp\left[\frac{-\gamma_b}{\sin^2\phi}\right] d\phi \tag{6-44}$$

AWGN信道中MPSK的误码率可类似写成下面的简单公式[13]

$$P_s = \frac{1}{\pi}\int_0^{(M-1)\pi/M} \exp\left[\frac{-g\gamma_s}{\sin^2\phi}\right] d\phi \tag{6-45}$$

其中$g = \sin^2(\pi/M)$。注意式（6-45）是准确式，不是对应于通式$\alpha_M \mathrm{Q}\left(\sqrt{\beta_M \gamma_s}\right)$的近似值。式（6-45）是有限区间上的积分，被积函数是简单的三角函数，所以很容易通过计算机或计算器进行数值计算。

182

6.3　衰落信道

在AWGN信道中，误码率的大小取决于接收信号的信噪比或者γ_s。在衰落信道中，阴影和多径衰落使接收信号的功率随空间和时间随机变化，使得γ_s是一个概率密度函数为$p_{\gamma_s}(\gamma)$的随机变量，因此，误码率$P_s(\gamma_s)$也是随机的。在此情况下，性能评价的指标取决于衰落信道变化的快慢，有以下3种：

- 中断率P_{out}，定义为γ_s低于某个给定值的概率，该给定值对应于可容忍的最大误码率。

• 平均错误率 $\bar{P}_s$，这是针对随机变量γ_s取平均的结果。

• 中断率和平均错误率的结合，定义为一定时间或空间内的平均错误率。

当衰落的相干时间与码元间隔可比（$T_s \approx T_c$）时，码元间隔内信号的衰落程度近似不变，这时常用平均误码率来衡量性能。少数的错误比特或者可以通过编码纠正，或者被丢弃或重传从而对端到端的性能无严重影响，因此平均错误率是这种情形下衡量信道质量的一个合理指标。

当信道衰落很慢（$T_s << T_c$）时，深衰落将会影响到很多连续的符号(simultaneous symbol)，从而引起一般编码无力纠正的突发错误。这样的突发错误将严重恶化端到端的性能。除非大幅度提高发射功率，否则便无法保证在任何时刻和任何地点都能达到对性能的最低要求。在这种情况下，可规定一定的中断率以允许少部分时间或地点内不能正常通信。当信道衰落是快慢结合时，例如对数正态的阴影和快变瑞利衰落同时存在的情形，可结合采用中断率和平均误码率。

另外当$T_c << T_s$时，衰落会被匹配滤波器平均掉。所以对于非常快的衰落，信道的性能与AWGN信道一样。

6.3.1 中断率

设γ_0是达到一定性能所必须的最小信噪比，中断率定义为

$$P_{\text{out}} = p(\gamma_s < \gamma_0) = \int_0^{\gamma_0} p_{\gamma_s}(\gamma)\,\mathrm{d}\gamma \tag{6-46}$$

例如，在数字话音业务中，误比特率为$P_b = 10^{-3}$时人耳基本上觉察不到误码带来的噪音。若对于端利衰落的BPSK信号，当$\gamma_b < 7\text{dB}$时就代表发生了中断，故设$\gamma_0 = 7\text{dB}$。

瑞利衰落时中断率为

$$P_{\text{out}} = \int_0^{\gamma_0} \frac{1}{\bar{\gamma}_s} \mathrm{e}^{-\gamma_s/\bar{\gamma}_s}\,\mathrm{d}\gamma_s = 1 - \mathrm{e}^{-\gamma_0/\bar{\gamma}_s} \tag{6-47}$$

求上式的逆可得到给定的中断率所需要的平均信噪比$\bar{\gamma}_s$为

$$\bar{\gamma}_s = \frac{\gamma_0}{-\ln(1 - P_{\text{out}})} \tag{6-48}$$

从分贝数来说，为了使性能可接受的时间百分比是$100 \times (1-P_{\text{out}})\%$，$10\log\gamma_s$必须比$10\log\gamma_0$高出$F_d = -10\log[-\ln(1-P_{\text{out}})]$分贝，一般称$F_d$为分贝衰落余量。

183

例6.4：对于慢变瑞利衰落信道中的BPSK调制，求$\bar{\gamma}_b$使得在95%的时间或空间内满足$P_b(\gamma_b) < 10^{-4}$。

解：在AWGN信道中，BPSK达到$P_b(\gamma_b) < 10^{-4}$的误比特率需要$\gamma_b = 8.5\text{dB}$。因为$P_b(\gamma_b) = Q\left(\sqrt{2\gamma_b}\right)$，$P_b(10^{0.85}) = 10^{-4}$。因此$\gamma_0 = 8.5\text{dB}$。为满足$P_{\text{out}} = p(\gamma_b < \gamma_0) = 0.05$，由式(6-48)可得

$$\bar{\gamma}_b = \frac{\gamma_0}{-\ln(1 - P_{\text{out}})} = \frac{10^{0.85}}{-\ln(1 - 0.05)} = 21.4\ \text{dB} \tag{6-49}$$

6.3.2 平均误码率

$T_s \approx T_c$时采用平均误码率作为性能指标。假设在一个符号周期内γ_s近似不变，则衰落信道中的平均误码率是将AWGN信道中的误码率按衰落分布进行积分

$$\bar{P}_s = \int_0^{\infty} P_s(\gamma) p_{\gamma_s}(\gamma)\,\mathrm{d}\gamma \tag{6-50}$$

其中$P_s(\gamma)$是AWGN信道中信噪比为γ时的误码率，可用表6-1中的近似式进行计算。若已知条件是衰落幅度r的分布（如给定瑞利衰落、莱斯衰落、对数正态衰落），可用变量替换的方法求出$p_{\gamma_s}(\gamma)$：

$$p_{\gamma_s}(\gamma)\,\mathrm{d}\gamma = p(r)\,\mathrm{d}r \tag{6-51}$$

例如，瑞利衰落中接收信号的幅度r服从瑞利分布

$$p(r) = \frac{r}{\sigma^2}\mathrm{e}^{-r^2/2\sigma^2} \quad r \geqslant 0 \tag{6-52}$$

信号功率服从均值为$2\sigma^2$的指数分布。给定幅度r时的符号信噪比为

$$\gamma = \frac{r^2 T_s}{2\sigma_n^2} \tag{6-53}$$

其中 $\sigma_n^2 = N_0/2$，它是同相和正交分量中噪声的方差[1]。对上式两边求导：

$$\mathrm{d}\gamma = \frac{rT_s}{\sigma_n^2}\,\mathrm{d}r \tag{6-54}$$

将式（6-53）和式（6-54）代入式（6-52）和式（6-51）可得

184

$$p_{\gamma_s}(\gamma) = \frac{\sigma_n^2}{\sigma^2 T_s}\mathrm{e}^{-\gamma\sigma_n^2/\sigma^2 T_s} \tag{6-55}$$

因为平均符号信噪比为 $\bar{\gamma}_s = \sigma^2 T_s/\sigma_n^2$，所以上式可写成

$$p_{\gamma_s}(\gamma) = \frac{1}{\bar{\gamma}_s}\mathrm{e}^{-\gamma/\bar{\gamma}_s} \tag{6-56}$$

这个结果是指数分布。对二元信号可写成

$$p_{\gamma_b}(\gamma) = \frac{1}{\bar{\gamma}_b}\mathrm{e}^{-\gamma/\bar{\gamma}_b} \tag{6-57}$$

将式（6-6）按照式（6-57）的分布进行积分可得到瑞利信道中BPSK的平均误码率为

$$\bar{P}_b = \frac{1}{2}\left[1 - \sqrt{\frac{\bar{\gamma}_b}{1+\bar{\gamma}_b}}\right] \approx \frac{1}{4\bar{\gamma}_b} \quad \text{(BPSK)} \tag{6-58}$$

式中的近似针对较大的$\bar{\gamma}_b$。用式（6-57）对式（6-31）做类似的平均可得到2FSK在瑞

1. 原书为PSD（功率谱密度），疑有误，此处已改。——译者注

利衰落中的平均误码率为

$$\bar{P}_b=\frac{1}{2}\left[1-\sqrt{\frac{\bar{\gamma}_b}{2+\bar{\gamma}_b}}\right]\approx\frac{1}{4\bar{\gamma}_b}\quad(\text{二进制 FSK})\tag{6-59}$$

当信噪比较大时，BPSK和2FSK的性能趋向相同。对于非相干调制，假设在符号周期内信道相移基本不变，平均误码率可以类似求得。DPSK的平均误比特率是

$$\bar{P}_b=\frac{1}{2(1+\bar{\gamma}_b)}\approx\frac{1}{2\bar{\gamma}_b}\quad(\text{DPSK})\tag{6-60}$$

其中的近似也只是针对较大的$\bar{\gamma}_b$。注意，$\bar{\gamma}_b$较大时，DPSK比BPSK有大约3dB的损失，又因为DPSK能纠正信道或接收机引入的慢相位改变，所以实际损失要比3dB小一些。作误码率计算时，并没有把它考虑在内。3dB功率损失同样出现在AWGN的差分检测中。

如果考虑通用的近似式 $P_s\approx\alpha_M Q\left(\sqrt{\beta_M\gamma_s}\right)$，那么瑞利信道中的平均误码率近似为

$$\bar{P}_s\approx\int_0^\infty\alpha_M Q\left(\sqrt{\beta_M\gamma}\right)\cdot\frac{1}{\bar{\gamma}_s}\mathrm{e}^{-\gamma/\bar{\gamma}_s}\,\mathrm{d}\gamma_s=\frac{\alpha_M}{2}\left[1-\sqrt{\frac{0.5\beta_M\bar{\gamma}_s}{1+0.5\beta_M\bar{\gamma}_s}}\right]\approx\frac{\alpha_M}{2\beta_M\bar{\gamma}_s}\tag{6-61}$$

上式中最后一个近似仅对较大的信噪比有效。 185

比较一下各种调制方式在AWGN信道和衰落信道中的误比特率，结果是很有意思的。二进制PSK、FSK和DPSK在AWGN中的误比特率随γ_b的增加而指数下降。但在衰落信道中，所有调制方式的误码率随着$\bar{\gamma}_b$的增加而线性下降。多进制调制中情形类似。因此，为了达到一个给定的误比特率P_b，特别是比较小的P_b，衰落信道需要的功率要比AWGN信道大得多。例如图6-1，图中画出了BPSK在AWGN信道和瑞利信道中的误码

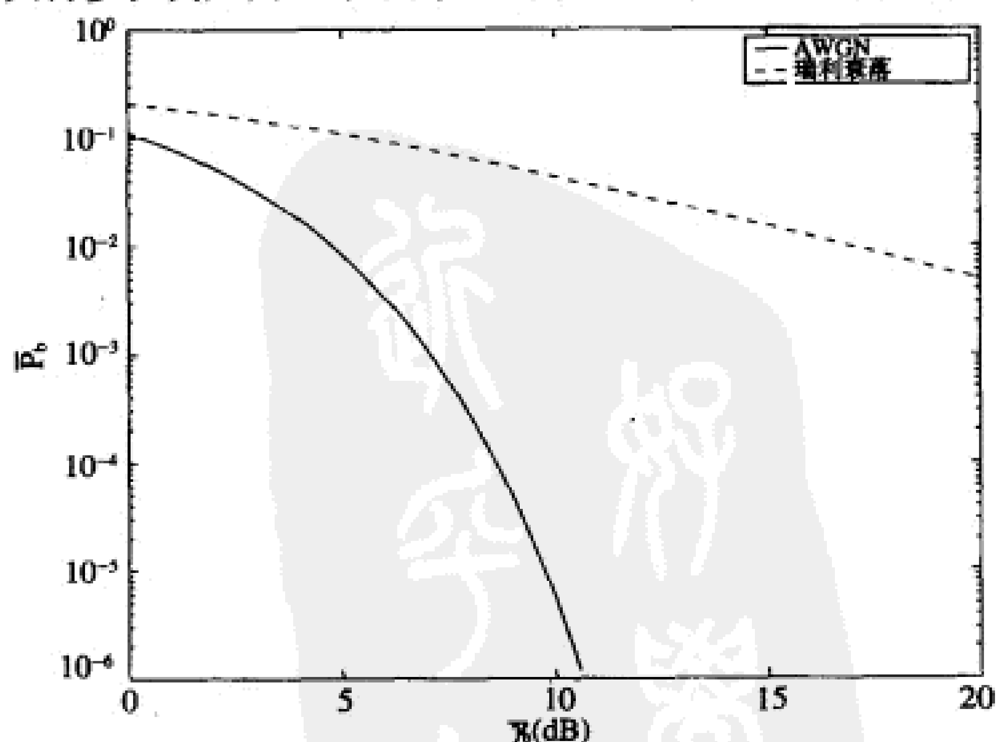


图6-1 AWGN信道和瑞利衰落信道中BPSK的平均误比特率

率。从图中可以看出，AWGN信道大约需要6.8dB的信噪比以保证10^{-3}的误比特率，而衰落信道大约需要24dB的信噪比才能保持相同的误比特率。对于MQAM，图6-2基于近似式（6-24）和式（6-61）画出了类似的误码率曲线。不难看出，要想降低对发射功率的要求，就必须要引入一些技术来克服衰落的影响。我们将会在以后的章节中讨论这类技术，包括分集合并、扩频和RAKE接收机等。

186

瑞利衰落是最坏的衰落之一。图6-3画出了不同参数的Nakagami衰落中BPSK的平均误比特率。随着参数m增大，衰落程度减小，平均误比特率趋向于AWGN信道的误比特率。

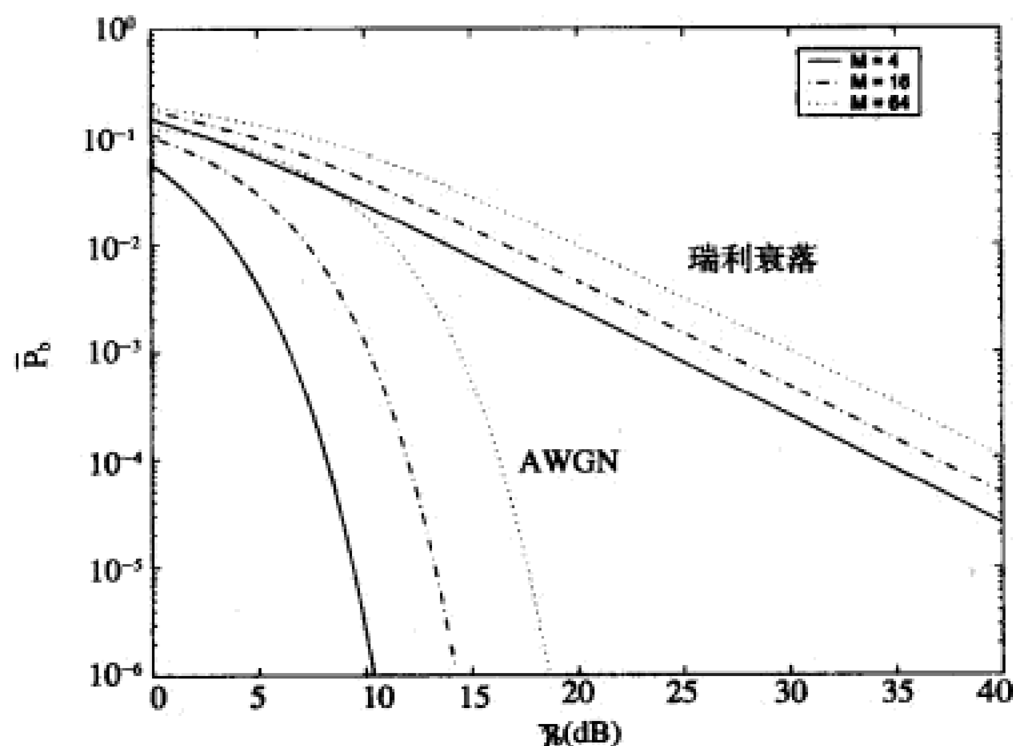


图6-2 AWGN信道和瑞利衰落信道中MQAM的平均误比特率

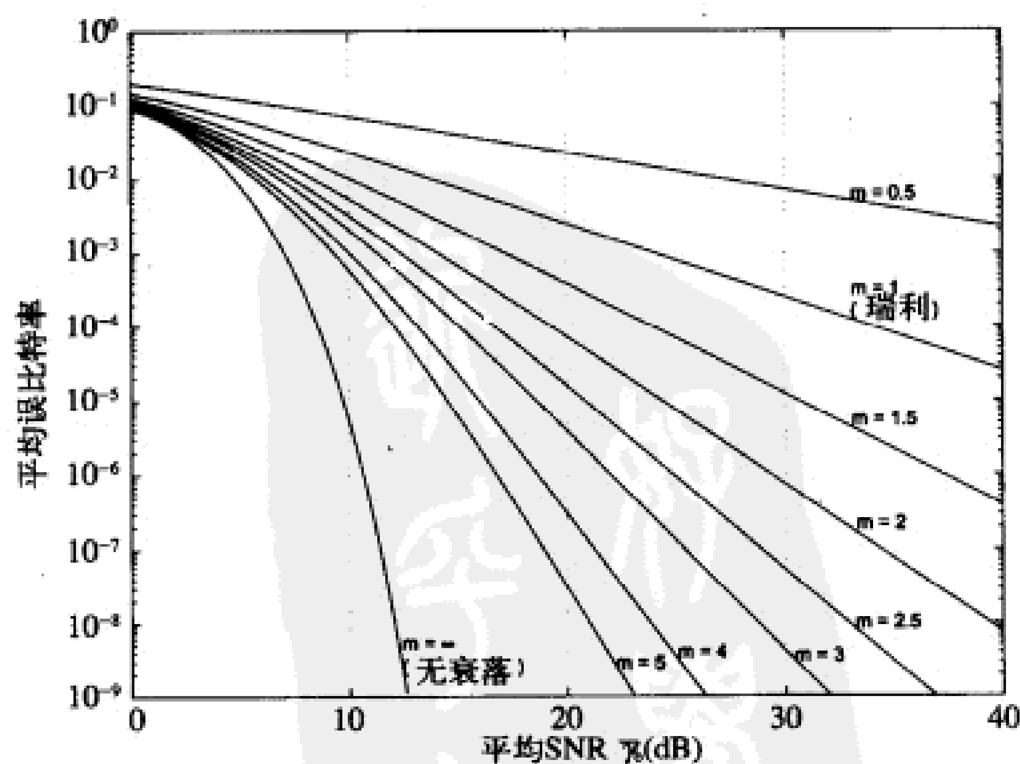


图6-3 Nakagami衰落信道中BPSK的平均误比特率

6.3.3 用矩母函数求平均误码率

在分析衰落信道中各种调制的性能时，**矩母函数**（moment generating function, GMF）是一个非常有用的工具，无论是分集或非分集的情形都能适用。本节将用矩母函数来研究衰落信道中的平均误码率，下一章将用它来研究分集系统的性能。

非负随机变量γ的矩母函数定义为：

$$\mathcal{M}_\gamma(s)=\int_0^\infty p_\gamma(\gamma)\mathrm{e}^{s\gamma}\,\mathrm{d}\gamma \tag{6-62}$$

其中$p_\gamma(\gamma)$，$\gamma\geqslant 0$是γ的概率密度函数。若将s取负号，上式就是$p_\gamma(\gamma)$的拉普拉斯变换，即$\mathcal{L}[p_\gamma(\gamma)]=\mathcal{M}_\gamma(-s)$。因此，大部分衰落分布的矩母函数都可以通过标准的拉普拉斯变换或数值积分的方法得到。文献[9]的5.1节给出了常见衰落分布的矩母函数，现列于下：

- 瑞利:

$$\mathcal{M}_{\gamma_s}(s)=(1-s\bar{\gamma}_s)^{-1} \tag{6-63}$$

- 莱斯（K因子）：

$$\mathcal{M}_{\gamma_s}(s)=\frac{1+K}{1+K-s\bar{\gamma}_s}\exp\left[\frac{Ks\bar{\gamma}_s}{1+K-s\bar{\gamma}_s}\right] \tag{6-64}$$

- Nakagami-m：

$$\mathcal{M}_{\gamma_s}(s)=\left(1-\frac{s\bar{\gamma}_s}{m}\right)^{-m} \tag{6-65}$$

187

顾名思义，矩母函数的名字说明它是矩之母，即可用矩母函数来求出各阶矩：

$$E[\gamma^n]=\frac{\partial^n}{\partial s^n}[\mathcal{M}_{\gamma_s}(s)]|_{s=0} \tag{6-66}$$

为了用矩母函数计算衰落信道的平均误码率，我们先把AWGN信道中的误码率表示成γ_s的指数函数

$$P_s=c_1\exp[-c_2\gamma_s] \tag{6-67}$$

其中c_1和c_2都是常量。或者把误码率表示成这种指数函数在有限区间上的积分：

$$P_s=\int_A^B c_1\exp[-c_2(x)\gamma_s]\,\mathrm{d}x \tag{6-68}$$

其中c_2可以和积分变量x有关，但它不能出现在积分限上，信噪比γ_s与积分变量无关并且也不出现在积分限上。这些表达形式使衰落信道中的平均误码率可以用矩母函数来表示。

具体来说，如果$P_s=c_1\exp(-c_2\gamma_s)$，那么

$$\bar{P}_s=\int_0^\infty c_1\exp[-c_2\gamma]p_{\gamma_s}(\gamma)\,\mathrm{d}\gamma=c_1\mathcal{M}_{\gamma_s}(-c_2) \tag{6-69}$$

例如对于DPSK，$c_1=1/2$，$c_2=1$，因此任意衰落的DPSK信号平均误比特率为

$$\bar{P}_b=\tfrac{1}{2}\mathcal{M}_{\gamma_s}(-1) \tag{6-70}$$

其中$M_{\gamma_s}(s)$为衰落分布的矩母函数。对瑞利衰落，矩母函数是式（6-63），$s=-1$，于

是$\bar{P}_b=[2(1+\bar{\gamma}_b)]^{-1}$，与式（6-60）的结果一致。

如果P_s是式（6-68）的积分形式，那么

$$\begin{aligned}\bar{P}_s&=\int_0^\infty\int_A^B c_1\exp[-c_2(x)\gamma]\,\mathrm{d}x\,p_{\gamma_s}(\gamma)\,\mathrm{d}\gamma\\&=c_1\int_A^B\left[\int_0^\infty\exp[-c_2(x)\gamma]p_{\gamma_s}(\gamma)\,\mathrm{d}\gamma\right]\mathrm{d}x\\&=c_1\int_A^B\mathcal{M}_{\gamma_s}(-c_2(x))\,\mathrm{d}x\end{aligned}\tag{6-71}$$

此时的平均误码率是对矩母函数在有限区间上的积分，往往能得到这个积分的闭式解，或者比较容易得到数值解。

现在我们针对具体的调制方式和衰落信道用矩母函数来求解。式（6-33）给出了AWGN信道相干解调的通用误码率近似式，这个近似式是用高斯Q函数表示的。对式（6-33）中的符号稍做变化，令$\alpha=\alpha_M$、$g=\beta_M/2$，则

188

$$P_s(\gamma_s)=\alpha\mathrm{Q}\left(\sqrt{2g\gamma_s}\right)\tag{6-72}$$

α和g是取决于所用调制方式的常量。这样的符号变化是为了便于使用矩母函数得到误码率的准确表达式。将Q函数的等效表示形式（6-43）代入得

$$P_s=\frac{\alpha}{\pi}\int_0^{\pi/2}\exp\left[\frac{-g\gamma}{\sin^2\phi}\right]\mathrm{d}\phi\tag{6-73}$$

上式正好具有式（6-68）的形式。这样，若某调制方式在AWGN信道中的误码率为$P_s=\alpha\mathrm{Q}\left(\sqrt{2g\gamma_s}\right)$，则它在衰落信道中的平均误码率为

$$\begin{aligned}\bar{P}_s&=\frac{\alpha}{\pi}\int_0^\infty\int_0^{\pi/2}\exp\left[\frac{-g\gamma}{\sin^2\phi}\right]\mathrm{d}\phi\,p_{\gamma_s}(\gamma)\,\mathrm{d}\gamma\\&=\frac{\alpha}{\pi}\int_0^{\pi/2}\left[\int_0^\infty\exp\left[\frac{-g\gamma}{\sin^2\phi}\right]p_{\gamma_s}(\gamma)\,\mathrm{d}\gamma\right]\mathrm{d}\phi\\&=\frac{\alpha}{\pi}\int_0^{\pi/2}\mathcal{M}_{\gamma_s}\left(\frac{-g}{\sin^2\phi}\right)\mathrm{d}\phi\end{aligned}\tag{6-74}$$

其中$\mathcal{M}_{\gamma_s}(s)$是概率密度函数$P_{\gamma_s}(\gamma)$的矩母函数。表6-1中给出了信道中$P_s\approx\alpha\mathrm{Q}\left(\sqrt{2g\gamma_s}\right)$形式的误码率近似式，因此式（6-74）给出了这些调制方式在衰落信道中的平均误码率的近似值。此外，对于相干解调的MPSK，注意到式（6-45）给出的误码率已经具有式（6-68）的形式，所以可以得到以式（6-74）形式表示的MPSK的准确平均误码率

$$\begin{aligned}\bar{P}_s&=\int_0^\infty\frac{1}{\pi}\int_0^{(M-1)\pi/M}\exp\left[\frac{-g\gamma_s}{\sin^2\phi}\right]\mathrm{d}\phi\,p_{\gamma_s}(\gamma)\,\mathrm{d}\gamma\\&=\frac{1}{\pi}\int_0^{(M-1)\pi/M}\left[\int_0^\infty\exp\left[\frac{-g\gamma_s}{\sin^2\phi}\right]p_{\gamma_s}(\gamma)\,\mathrm{d}\gamma\right]\mathrm{d}\phi\\&=\frac{1}{\pi}\int_0^{(M-1)\pi/M}\mathcal{M}_{\gamma_s}\left(-\frac{g}{\sin^2\phi}\right)\mathrm{d}\phi\end{aligned}\tag{6-75}$$

其中$g=\sin^2(\pi/M)$取决于MPSK星座图的大小。瑞利、莱斯和Nakagami-m分布的矩母函

数已分别由式（6-63）、式（6-64）和式（6-65）给出。将$s=-g/\sin^2(\phi)$代入可得

• 瑞利衰落信道：

$$\mathcal{M}_{\gamma_s}\left(-\frac{g}{\sin^2\phi}\right)=\left(1+\frac{g\bar{\gamma}_s}{\sin^2\phi}\right)^{-1} \tag{6-76}$$

• 因子为k的莱斯分布：

$$\mathcal{M}_{\gamma_s}\left(-\frac{g}{\sin^2\phi}\right)=\frac{(1+K)\sin^2\phi}{(1+K)\sin^2\phi+g\bar{\gamma}_s}\exp\left[-\frac{Kg\bar{\gamma}_s}{(1+K)\sin^2\phi+g\bar{\gamma}_s}\right] \tag{6-77}$$

• 参数为m的Nakagami分布：

$$\mathcal{M}_{\gamma_s}\left(-\frac{g}{\sin^2\phi}\right)=\left(1+\frac{g\bar{\gamma}_s}{m\sin^2\phi}\right)^{-m} \tag{6-78}$$

189

这些都是简单的三角函数，因此容易求得式（6-74）和式（6-75）中的积分。

例6.5：用矩母函数的方法求出Nakagami信道中BPSK的平均误码率表达式。

解：由于在AWGN信道中，BPSK的误比特率为$P_b=\mathrm{Q}\left(\sqrt{2\gamma_b}\right)$，所以式（6-72）中的$\alpha=1$、$g=1$。Nakagami-m分布的矩母函数由式（6-78）给出。将这些代入式（6-74）得

$$\bar{P}_b=\frac{1}{\pi}\int_0^{\pi/2}\left(1+\frac{\bar{\gamma}_b}{m\sin^2\phi}\right)^{-m}\mathrm{d}\phi$$

MQAM在AWGN信道中的误码率的准确表达式（6-23）不但包含Q函数，还包含它的平方。$\mathrm{Q}^2(z)$也有一种等效表达式，因此可以将前述方法推广到MQAM。文献[15]给出$\mathrm{Q}^2(z)$的等效表达式为

$$\mathrm{Q}^2(z)=\frac{1}{\pi}\int_0^{\pi/4}\exp\left[\frac{-z^2}{2\sin^2\phi}\right]\mathrm{d}\phi \tag{6-79}$$

可以发现上式和Q函数的等效表达式（6-43）基本一致，差别只是积分上限从$\pi/2$变成了$\pi/4$。把式（6-23）用$\mathrm{Q}(z)$和$\mathrm{Q}^2(z)$的等效表达式写成

$$\begin{aligned}P_s(\gamma_s)=&\frac{4}{\pi}\left(1-\frac{1}{\sqrt{M}}\right)\int_0^{\pi/2}\exp\left(-\frac{g\gamma_s}{\sin^2\phi}\right)\mathrm{d}\phi\\&-\frac{4}{\pi}\left(1-\frac{1}{\sqrt{M}}\right)^2\int_0^{\pi/4}\exp\left[-\frac{g\gamma_s}{\sin^2\phi}\right]\mathrm{d}\phi\end{aligned} \tag{6-80}$$

其中$g=1.5/(M-1)$。于是，MQAM在衰落信道中的平均误码率为

$$\begin{aligned}\bar{P}_s=&\int_0^\infty P_s(\gamma)p_{\gamma_s}(\gamma)\,\mathrm{d}\gamma\\=&\frac{4}{\pi}\left(1-\frac{1}{\sqrt{M}}\right)\int_0^{\pi/2}\int_0^\infty\exp\left[-\frac{g\gamma}{\sin^2\phi}\right]p_{\gamma_s}(\gamma)\,\mathrm{d}\gamma\,\mathrm{d}\phi\\&-\frac{4}{\pi}\left(1-\frac{1}{\sqrt{M}}\right)^2\int_0^{\pi/4}\int_0^\infty\exp\left[-\frac{g\gamma}{\sin^2\phi}\right]p_{\gamma_s}(\gamma)\,\mathrm{d}\gamma\,\mathrm{d}\phi\\=&\frac{4}{\pi}\left(1-\frac{1}{\sqrt{M}}\right)\int_0^{\pi/2}\mathcal{M}_{\gamma_s}\left(-\frac{g}{\sin^2\phi}\right)\mathrm{d}\phi\\&-\frac{4}{\pi}\left(1-\frac{1}{\sqrt{M}}\right)^2\int_0^{\pi/4}\mathcal{M}_{\gamma_s}\left(-\frac{g}{\sin^2\phi}\right)\mathrm{d}\phi\end{aligned} \tag{6-81}$$

190

可以看出，只要计算两个对矩母函数的有限积分就可以得到平均误码率的准确值。这两个积分一般可以得到闭式解，或者比较容易得到数值解。

这种矩母函数的方法还可用于非相干及差分的情形。例如，非相干MFSK在AWGN信道中的误码率由式（6-32）给出，它是有限项的和，每一项都具有式（6-67）的形式。于是，非相干MFSK在衰落信道中的平均误码率为

$$\begin{aligned}\bar{P}_s &= \int_0^\infty \sum_{m=1}^{M}(-1)^{m+1}\binom{M-1}{m}\frac{1}{m+1}\exp\left[\frac{-m\gamma}{m+1}\right]p_{\gamma_s}(\gamma)\,\mathrm{d}\gamma \\ &= \sum_{m=1}^{M}(-1)^{m+1}\binom{M-1}{m}\frac{1}{m+1}\left[\int_0^\infty \exp\left[\frac{-m\gamma}{m+1}\right]p_{\gamma_s}(\gamma)\,\mathrm{d}\gamma\right] \\ &= \sum_{m=1}^{M}(-1)^{m+1}\binom{M-1}{m}\frac{1}{m+1}\mathcal{M}_{\gamma_s}\left(-\frac{m}{m+1}\right)\end{aligned} \tag{6-82}$$

对于差分MPSK，文献[16]给出其平均误码率为

$$P_s = \frac{\sqrt{g}}{2\pi}\int_{-\pi/2}^{\pi/2}\frac{\exp\left[-\gamma_s\left(1-\sqrt{1-g}\cos\theta\right)\right]}{1-\sqrt{1-g}\cos\theta}\,\mathrm{d}\theta \tag{6-83}$$

其中$g=\sin^2(\pi/M)$。上式具有式（6-68）的形式，于是可用矩母函数写出平均误码率为

$$\bar{P}_s = \frac{\sqrt{g}}{2\pi}\int_{-\pi/2}^{\pi/2}\frac{\mathcal{M}_{\gamma_s}\left(-\left(1-\sqrt{1-g}\cos\theta\right)\right)}{1-\sqrt{1-g}\cos\theta}\,\mathrm{d}\theta \tag{6-84}$$

更多关于用矩母函数求解各种调制在各种衰落信道中的平均误码率的问题，请参考文献[9]的8.2节。

6.3.4 中断率和平均误码率的结合

信道可能会同时具有慢衰落和快衰落，比如同时存在对数正态阴影衰落和瑞利衰落。此时常用的性能指标是中断率和平均误码率的结合。当慢衰落低于某个目标值时发生中断，非中断时的平均性能是通过对快衰落进行平均求得。我们使用如下记号：

- $\bar{\bar{\gamma}}_s$表示路径损耗为定值时，对阴影和快衰落求平均得到的符号信噪比。
- $\bar{\gamma}_s$表示路径损耗为定值、阴影衰落为定值时，对快衰落求平均得到的符号信噪比。因为阴影衰落值是随机给定的，所以$\bar{\gamma}_s$是随机的，它对阴影求平均便是$\bar{\bar{\gamma}}_s$。
- γ_s表示路径损耗为定值、阴影衰落为定值、快衰落也为定值时的符号信噪比。因为阴影和快衰落是随机的，所以γ_s是随机的。γ_s只对快衰求平均便是$\bar{\gamma}_s$，γ_s同时对快衰和阴影求平均便是$\bar{\bar{\gamma}}_s$。

191

然后我们就可以定义出依概率$1-P_{\text{out}}$的平均错误概率$\bar{P}_s$。对于某个目标值$\bar{\gamma}_{s0}$，称事件$\bar{\gamma}_s<\bar{\gamma}_{s0}$为中断，无中断（即$\bar{\gamma}_s>\bar{\gamma}_{s0}$）时的平均误码率定义为给定$\bar{\gamma}_s$条件下对快衰落取平均得到的平均误码率

$$\bar{P}_s = \int_0^\infty P_s(\gamma_s)p(\gamma_s \mid \bar{\gamma}_s)\,\mathrm{d}\gamma_s \tag{6-85}$$

目标值 $\bar{\gamma}_{s0}$ 由最大可接受的平均误码率 $\bar{P}_{s0}$ 决定，它满足

$$\bar{P}_{s0}=\int_0^{\infty} P_s(\gamma_s)p(\gamma_s \mid \bar{\gamma}_{s0})\,\mathrm{d}\gamma_s \tag{6-86}$$

显然，只要 $\bar{\gamma}_s>\bar{\gamma}_{s0}$ 就有 $P_s<\bar{P}_{s0}$。

例6.6：某信道同时具有对数正态阴影衰落和瑞利快衰落，阴影衰落的标准差 $\sigma_{\psi_{dB}}=8\text{dB}$。假设采用BPSK调制，最大可接受的平均误码率为 $\bar{P}_{b0}=10^{-4}$，对应要求 $\bar{\gamma}_{b0}=34\text{dB}$。要求以 $1-P_{out}=0.95$ 的概率保证 $\bar{P}_b\leqslant10^{-4}$。求 $\bar{\bar{\gamma}}_b$。

解：假设采用分贝单位，则 $(\bar{\gamma}_b-\bar{\bar{\gamma}}_b)/\sigma_{\psi_{dB}}$ 服从标准正态分布，由 $p(\bar{\gamma}_b>\bar{\gamma}_{b0})=1-P_{out}$ 得：

$$p(\bar{\gamma}_b>34)=p\left(\frac{\bar{\gamma}_b-\bar{\bar{\gamma}}_b}{\sigma_{\psi_{dB}}}\geqslant\frac{34-\bar{\bar{\gamma}}_b}{\sigma_{\psi_{dB}}}\right)=Q\left(\frac{34-\bar{\bar{\gamma}}_b}{\sigma_{\psi_{dB}}}\right)=1-P_{out} \tag{6-87}$$

通过查Q函数表，或者借助计算机可得 $(34-\bar{\bar{\gamma}}_b)/8=-1.6$，即 $\bar{\bar{\gamma}}_b=46.8\text{dB}$。

6.4 多普勒频移

多普勒频移会使差分检测出现背景误码。因为差分调制用前一个符号的相位作为当前符号的参考，如果经过一个符号周期后信道相移去相关，那么这个参考就不再正确了，这将导致误码，即使加大发送功率也一样。相邻符号间信道相位的相关程度是多普勒频移 $f_D=v/\lambda$ 和码元间隔 T_s 的函数，因此快衰落引起的性能恶化程度与 f_D 和 T_s 有关。

192

Bello和Nelin[17]最早研究了多普勒频移引起的背景误码问题，并针对高斯多普勒功率谱给出了非相干FSK和DPSK中背景误码率的解析式。但他们给出的表达式没有闭式解，只能用数值方法来计算。对于快变莱斯信道，可以用矩母函数的方法得到DPSK误比特率的闭式解，其矩母函数可通过复高斯随机变量的广义二次型得到，详见文献[1]的附录B及文献[12]的附录B。也可以用Q函数的等效表达式来求得（详见文献[9]的8.2.5节），这两种方法求得的DPSK平均误比特率为

$$\bar{P}_b=\frac{1}{2}\left(\frac{1+K+\bar{\gamma}_b(1-\rho_C)}{1+K+\bar{\gamma}_b}\right)\exp\left[-\frac{K\bar{\gamma}_b}{1+K+\bar{\gamma}_b}\right] \tag{6-88}$$

其中 ρ_C 是经过 T_b 时间的信道相关系数，K 是莱斯因子，$\bar{\gamma}_b$ 是平均每比特的信噪比。对瑞利分布（$K=0$）上式简化为

$$\bar{P}_b=\frac{1}{2}\left(\frac{1+\bar{\gamma}_b(1-\rho_C)}{1+\bar{\gamma}_b}\right) \tag{6-89}$$

令式（6-88）中的 $\bar{\gamma}_b\to\infty$，则背景误码率为

$$\bar{P}_{floor}=\frac{(1-\rho_C)\mathrm{e}^{-K}}{2}\quad(\text{DPSK}) \tag{6-90}$$

文献[18]用类似的方法得到了快变莱斯信道中DQPSK误比特率的界：

$$\bar{P}_b\leqslant\frac{1}{2}\left(1-\sqrt{\frac{(\rho_C\bar{\gamma}_s/\sqrt{2})^2}{(\bar{\gamma}_s+1)^2-(\rho_C\bar{\gamma}_s/\sqrt{2})^2}}\right)\exp\left[-\frac{(2-\sqrt{2})K\bar{\gamma}_s/2}{(\bar{\gamma}_s+1)-(\rho_C\bar{\gamma}_s/\sqrt{2})}\right] \tag{6-91}$$

PDG

其中K是莱斯因子，ρ_C是间隔T_s的信道相关系数，$\bar{\gamma}_s$是平均每符号的信噪比。令式中的$\bar{\gamma}_s\to\infty$，可得背景误码率为

193

$$\bar{P}_{\text{floor}}=\frac{1}{2}\left(1-\sqrt{\frac{(\rho_C/\sqrt{2})^2}{1-(\rho_C/\sqrt{2})^2}}\right)\exp\left[-\frac{(2-\sqrt{2})(K/2)}{1-\rho_C/\sqrt{2}}\right]\quad(\text{DQPSK})\tag{6-92}$$

由3.2.1节可知：信道的相关函数$A_C(t)$等于多普勒谱$S_C(f)$的傅里叶反变换。因此相关系数$\rho_C=A_C(T)/A_C(0)$，其中对DQPSK，$T=T_s$，对DPSK，$T=T_b$。文献[19]中表6-2给出了几种不同多普勒功率谱模型对应的ρ_C，其中B_D是信道的多普勒扩展。对于均匀散射的瑞利衰落，$\rho_C=J_0(2\pi f_D T_b)$，$K=0$，由式（6-90）可得DPSK的背景误码率为

$$\bar{P}_{\text{floor}}=\frac{1-J_0(2\pi f_D T_b)}{2}\approx 0.5(\pi f_D T_b)^2\tag{6-93}$$

其中$B_D=f_D=v/\lambda$是信道的最大多普勒频移。注意上式中背景误码随着$R=1/T_b$的增大而减小。这是一种普遍现象，因为速率增加时，符号周期会减小，符号周期间隔的信道相关性会增大。在数字通信中，性能随传输速率的增加而改善的现象并不多见，这是一例。

表6-2　不同多普勒功率谱模型及其相关系数

类　型	多普勒功率谱$S_C(f)$	$\rho_C=A_C(T)/A_C(0)$
矩形	$P_0/2B_D$, $\lvert f\rvert<B_D$	$\text{sinc}(2B_DT)$
高斯	$(P_0/\sqrt{\pi}B_D)e^{-f^2/B_D^2}$	$e^{-(\pi B_DT)^2}$
均匀散射	$P_0/\pi\sqrt{B_D^2-f^2}$, $\lvert f\rvert<B_D$	$J_0(2\pi B_DT)$
一阶巴特沃思	$P_0B_D/\pi(f^2+B_D^2)$	$e^{-2\pi B_DT}$

图6-4根据式（6-88）画出了在均匀散射的莱斯快衰落信道中，不同f_DT_b值时DPSK

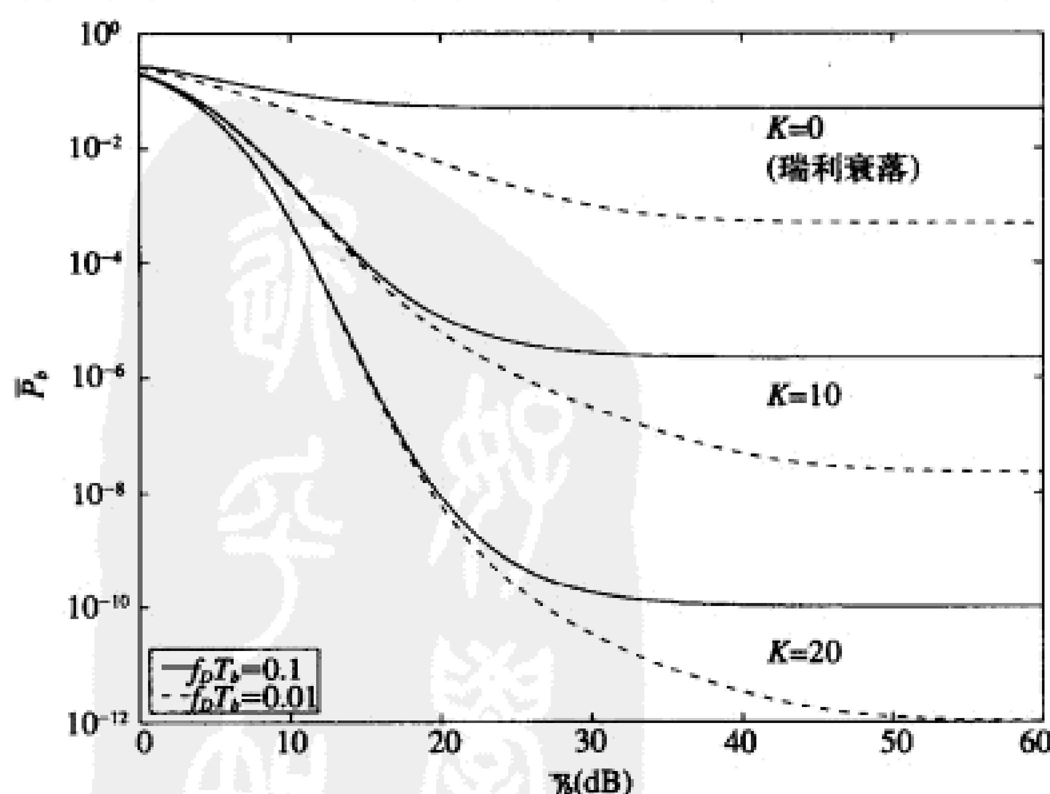


图6-4　均匀散射快变莱斯信道中DPSK的平均误比特率

的误码率曲线。在瑞利信道（$K=0$）中，$\bar{\gamma}_b=15\text{dB}$时背景误码开始起主导作用。随着$K$的增大，背景误码开始起主导作用所对应的$\bar{\gamma}_b$也增大。还可注意到，数据速率$R_b=1/T_b$增加一个数量级，背景误码大约会减小两个数量级。

例6.7：考虑一个均匀散射的瑞利信道，最大多普勒频移为$f_D=80\text{Hz}$。欲使DPSK的背景误码率小于10^{-4}，数据速率大致应该在什么范围内？

解：$\overline{P}_{\text{floor}}\approx 0.5(\pi f_D T_b)^2<10^{-4}$，代入$f_D=80\text{Hz}$可得

$$T_b<\frac{\sqrt{2\cdot 10^{-4}}}{\pi\cdot 80}=5.63\times 10^{-5}$$

由此可得所求数据速率范围是$R>17.77\text{kbit/s}$。

194

对于更为复杂的调制，很难通过解析的方法得到背景误码率，常用仿真来求解。文献[20~21]给出了采用了平方根升余弦滚降滤波器的π/4−DQPSK的背景误码率的仿真结果，这是IS-136 TDMA标准采用的调制方式。其结果表明背景误码率在10^{-3}和10^{-4}之间，且背景误码率随车速升高而升高，这也是因为车速越高时信道越不相关。

6.5 码间干扰

频率选择性衰落会导致码间干扰，其他时刻的码元通过不同时延的多径传播对当前时刻的码元形成干扰。提高发送功率同样提高了码间干扰的强度，所以码间干扰也会造成背景误码。码间干扰造成的背景误码和调制方式及码间干扰的特性有关，而码间干扰的特性又和信道特性及发送的符号序列有关，所以这一问题很难用解析的方法进行分析。

Bello和Nelin[22]最早研究了由码间干扰带所导致的误码率降低的问题，并针对高斯功率时延谱，导出了相干FSK和非相干DPSK的背景误码率的表达式。为了简化分析，文献[22]只考虑了相邻符号引起的码间干扰。即使在这样简单的条件下，所得结果还是非常复杂，只能近似估算。背景误码率还可以通过最坏发送序列给出，或者通过对所有可能发送序列求平均来给出，详见文献[23]的8.2节。这样得到的结果也非常复杂，因为它依赖于信道和所发送序列的特性。一种近似的方法是把码间干扰看成独立高斯噪声[24]，它使信噪比变为

$$\hat{\gamma}_s=\frac{P_r}{N_0B+I} \tag{6-94}$$

式中P_r是接收信号中直射部分的功率，I是码间干扰的功率。若信道是静态的，则误码率$P_s(\hat{\gamma}_s)$可按AWGN信道计算。若直射分量及码间干扰都经历了平衰落，那么$\hat{\gamma}_s$将是一个随机变量，设其分布为$p(\hat{\gamma}_s)$，则平均误码率为$\overline{P}_s=\int P_s(\hat{\gamma}_s)p(\hat{\gamma}_s)d\hat{\gamma}_s$。注意此时$\hat{\gamma}_s$是接收的功率$P_r$和码间干扰强度$I$这两个随机变量的比值，因此一般很难求得$p(\hat{\gamma}_s)$的闭式解。

195

背景误码率可以通过仿真来获得。仿真易于将各种不同的信道模型、调制方式及符号序列都考虑进来[20-21],[24-26]。其中Chuang[25]的工作广泛包括了多种调制方式（BPSK、DPSK、QPSK、OQPSK和MSK）、多种脉冲成形、多种功率时延谱（高斯、指数、两径等增益以及经验谱）。其结果表明：均方根时延扩展对背景误码率的影响比功率时延

PDG

谱的形状更大；脉冲成形对背景误码率的影响也很大，升余弦脉冲的滚降系数（见5.5节）β从0增到1时，背景误码率相应下降了一个数量级。

图6-5是Chuang的仿真结果之一。图中画出了BPSK、QPSK、OQPSK和MSK的背景误码率和归一化均方根时延扩展$d=\sigma_{T_m}/T_s$的关系，假设条件是静态信道的高斯功率时延谱。对于图中的各种调制方式，背景误码率的上界是$P_{floor}\leqslant d^2$，$0.02\leqslant d\leqslant 0.1$。其他文献如[24]的仿真结果也支持这个界。这个界的意义是：即便可接受的误码率可以达到10^{-2}，码间干扰也将使数据速率受到严重的限制。举例来说，典型城市环境中的均方根时延扩展大约是$\sigma_{T_m}=2.5\mu s$，按10^{-2}的背景误码率，需$d\leqslant 0.1$，即$\sigma_{T_m}<0.1T_s$，即要求传输速率$1/T_s\leqslant 0.1/\sigma_{T_m}=40$kbaud，这个速率一般是不能满足高速数据业务的要求的。乡村环境不像城市环境那样能将多径成分显著衰减，因此有更大的时延扩展，$\sigma_{T_m}\approx 25\mu s$，最大数据速率相应降低到4kbaud。

196

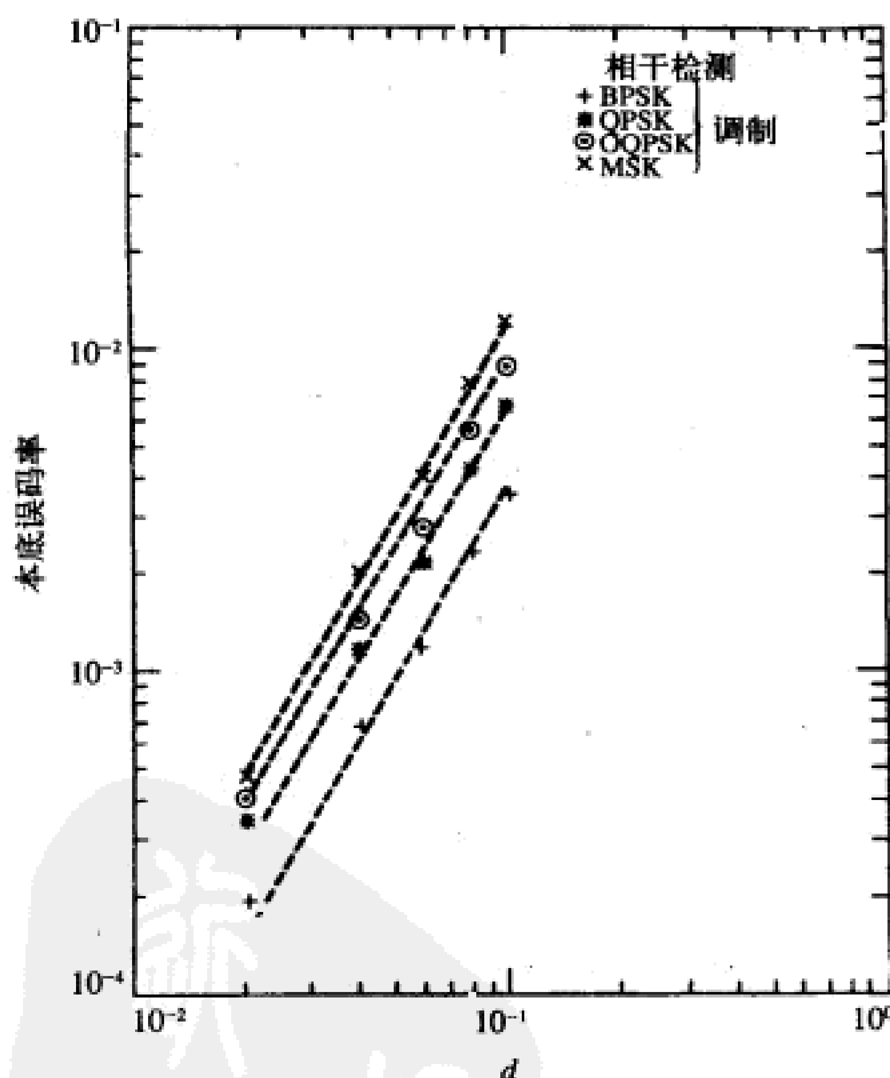


图6-5　背景误码率和高斯功率时延谱的归一化均方根时延扩展的关系

例6.8：假设信道的时延扩展为$\sigma_{T_m}=3\mu s$，请用近似式$\overline{P}_{floor}\leqslant(\sigma_{T_m}/T_s)^2$分别求出BPSK和QPSK调制的最大数据速率，要求误比特率小于10^{-3}。

解：对于BPSK，$\overline{P}_{floor}\leqslant(\sigma_{T_m}/T_b)^2$，因此$T_b\geqslant\sigma_{T_m}/\sqrt{\overline{P}_{floor}}=94.87\mu s$，传输速率为$R=1/T_b=10.54$kbit/s。对于QPSK同样可算出$T_s\geqslant\sigma_{T_m}/\sqrt{\overline{P}_{floor}}=94.87\mu s$，因为一个符号包含两个比特，所以数据速率为$R=2/T_s=21.01$kbit/s。这说明对于给定的数据速率，QPSK比BPSK更能抵抗

码间干扰，原因是QPSK的符号周期比BPSK要大。本题中若不用近似式$\overline{P}_{floor} \leqslant (\sigma_{T_m}/T_s)^2$，而是用图6-5中更为精确的曲线来计算，结论不变。

习题

6.1 某系统的数据速率为100bit/s。

(a) 如果用sinc脉冲来传输，且信道的带宽为10kHz，求符号的持续时间。

(b) 如果接收信号的信噪比为10dB，求采用4-QAM时的每符号的信噪比和每比特的信噪比。

(c) 在上一问中如果不用4-QAM而用16-QAM，求相应的每符号的信噪比和每比特的信噪比，并与4-QAM的结果作比较。

6.2 某BPSK调制中发送0和1的先验概率不相等，$p[s_n=0]=0.3$，$p[s_n=1]=0.7$。

(a) 在AWGN信道中，若将1调制为$s_1(t)=A\cos(2\pi f_c t)$，将0调制为$s_2(t)=-A\cos(2\pi f_c t)$，$A>0$。接收机结构如图5-17所示。求此时的误比特率$P_b$。

(b) 如果5.17所示接收机中的判决门限值可以改变，求能使错误率对称的门限值，即该门限满足：$p(\hat{m}=0|m=1)p(m=1)=p(\hat{m}=1|m=0)p(m=0)$。

(c) 现在我们把调制方式改为：1调制成$s_1(t)=A\cos(2\pi f_c t)$，0调制成$s_2(t)=-B\cos(2\pi f_c t)$。用图5-17中的接收机，判决门限为0。求满足$p(\hat{m}=0|m=1)p(m=1)=p(\hat{m}=1|\mathrm{m}=0)p(m=0)$的$A$、$B$的正值。

(d) 假设$E_b/N_0=10\text{dB}$，$N_0=0.1$，计算并比较以上三问中的P_b。哪种情况下P_b最小？

6.3 假设BPSK系统中发送信号为$s(t)=\pm g(t)\cos(2\pi f_c t)$，解调器的载波相对发送信号有相移$\phi$偏差，即图5-17中的恢复载波为$\cos(2\pi f_c t+\phi)$。求图5-17中能使误比特率最小的判决门限，并求出此最小的误比特率。

6.4 假设BPSK解调器中噪声叠加在积分器之后，如图6-6所示。图中的判决器当$\mathrm{Re}\{x\}\geqslant 0$时输出1，否则输出0。假设噪声为$n(t)=1.1e^{j\theta}$，其中$\theta$满足$p(\theta=n\pi/3)=1/6$，$n=0, 1, 2, 3, 4, 5$。假设发1对应$s(t)=A_c\cos(2\pi f_c t)$，发0对应$s(t)=-A_c\cos(2\pi f_c t)$，其中$A_c=\sqrt{2/T_b}=1$，发1和0的概率相等。求判决器出错的概率。

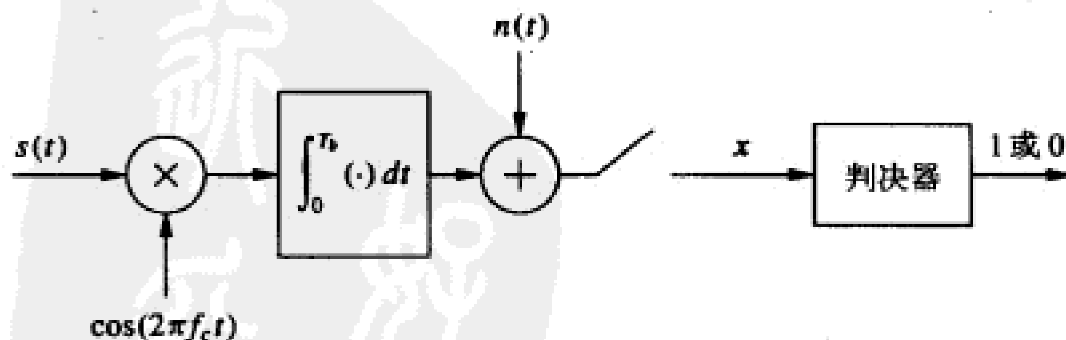


图6-6 题6.4的BPSK解调器

6.5 对于图6-7所示的各星座图，求误码率的近似值。

6.6 请在信噪比范围$0\leqslant\gamma_s\leqslant 30\text{dB}$内画出16QAM的准确误码率及表6-1中的近似误码率的曲线。并回答对于不同的γ_s，近似误差有何不同？为什么？

6.7 对于QPSK，请在信噪比范围$0\leqslant\gamma_s\leqslant 30\text{dB}$内画出用表6-1中公式算出的近似误码率和用Craig的公式算出的误码率准确值。并回答对于不同的γ_s，近似误差有何不同？为什么？

198

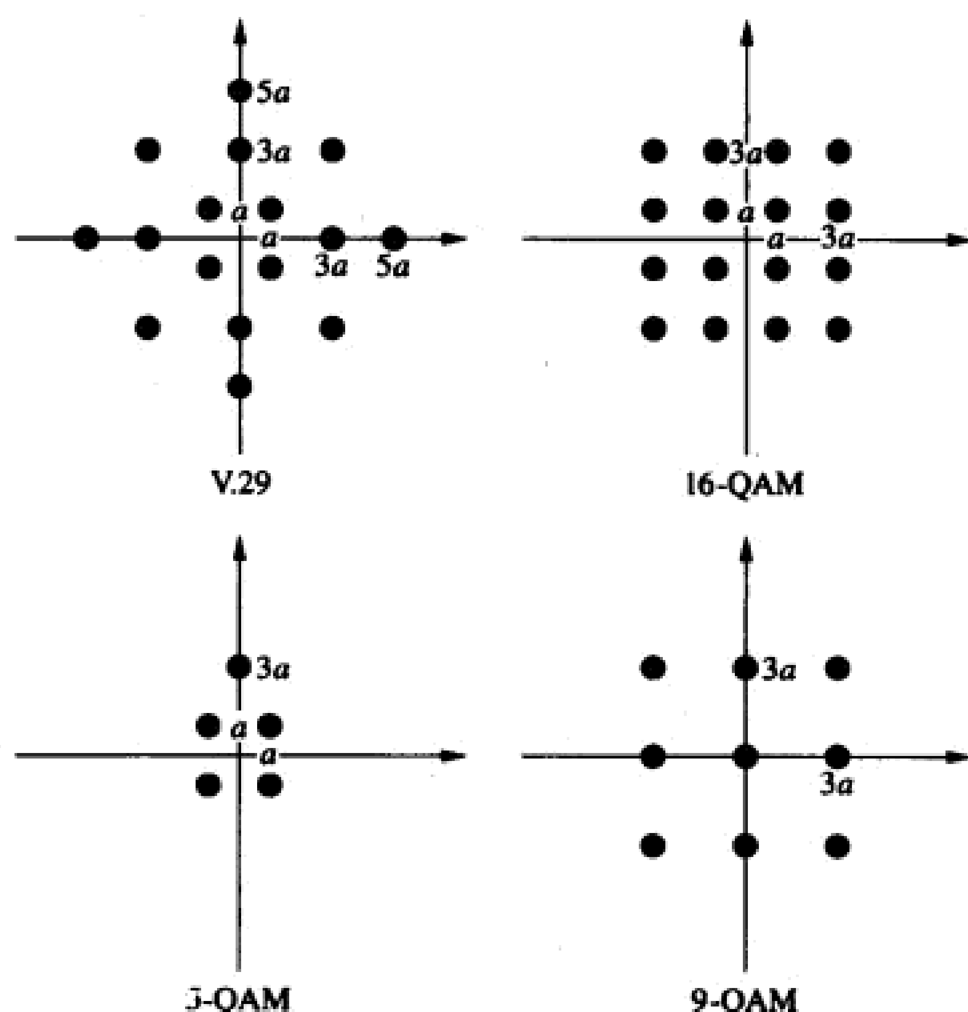


图6-7 题6.5的星座图

6.8 本题从Q函数的原始形式（6-42）推导出其等效形式（6-43）。为简单起见，考虑互补误差函数，最后再转换到Q函数。互补误差函数定义为

$$\operatorname{erfc}(x)=\frac{2}{\sqrt{\pi}}\int_x^{\infty}\mathrm{e}^{-t^2}\,\mathrm{d}t \tag{6-95}$$

和Q函数的等效形式（6-43）相对应，erfc(x)的等效形式为

$$\operatorname{erfc}(x)=\frac{2}{\pi}\int_0^{\pi/2}\mathrm{e}^{-x^2/\sin^2\theta}\,\mathrm{d}\theta \tag{6-96}$$

(a) 考虑积分

$$I_x(a)\triangleq\int_0^{\infty}\frac{\mathrm{e}^{-at^2}}{x^2+t^2}\,\mathrm{d}t \tag{6-97}$$

请证明$I_x(a)$满足如下的差分方程：

$$x^2 I_x(a)-\frac{\partial I_x(a)}{\partial a}=\frac{1}{2}\sqrt{\frac{\pi}{a}} \tag{6-98}$$

(b) 请解这个差分方程并导出

$$I_x(a) \triangleq \int_0^{\infty} \frac{e^{-at^2}}{x^2+t^2} dt = \frac{\pi}{2x} e^{ax^2} \operatorname{erfc}(x\sqrt{a}) \tag{6-99}$$

提示：$I_x(a)$是x和a的二元函数，解差分方程时可将x视为常量。

(c) 请在式（6-99）中置$a=1$，并对左边的积分作适当的变量代换，从而得到互补误差函数的等效表达式

$$\operatorname{erfc}(x) = \frac{2}{\pi} \int_0^{\pi/2} e^{-x^2/\sin^2\theta} d\theta$$

(d) 请将这个erfc的等效表达式进行转换得到Q函数的等效表达式。

6.9 某通信系统采用BPSK调制，平均信号功率为100W，接收端的噪声功率为4W。请问此系统的误码率是否适合于数据传输？是否适合于话音传输？若再假设存在平均信噪比$\bar{\gamma}_b = 20$dB的衰落，此条件下对刚才的两个问题的答案有无变化？

6.10 考虑一个900MHz的蜂窝系统，数据速率为64kbit/s，存在多径衰落。请问，当用户的速度分别为1英里/小时、10英里/小时和100英里/小时的时候，平均误码率和中断率两者中哪一个性能指标更合适？为什么？

6.11 请推导出瑞利衰落的信噪比的矩母函数表达式。

6.12 本题将说明，如果一个卫星系统必须要补偿阴影衰落的话，那么这个系统必然要走向破产。考虑一个轨道高度为500公里的LEO系统。假设自由空间传播落模型，没有多径衰落和阴影衰落。载波频率为900MHz，带宽为10kHz。手持终端的噪声功率谱密度为10^{-16}mW/Hz（总噪声功率为N_0B）。假设发送端和接收端均采用全向天线（天线增益为0dB）。在地球上一个半径为100km的圆内，此卫星系统必须要保证所有用户的误比特率不大于10^{-6}。

(a) 假设DPSK调制，求能够满足此误比特率要求的发射功率。

(b) 假设存在标准差为$\sigma_{\psi_{dB}} = 8$dB的对数正态阴影衰落。为了使此区域中90%的用户达到的误比特率为10^{-6}，重新求发射功率。

6.13 本题探讨从BPSK提升至高阶调制16PSK时的功率损失。

(a) 以能量E_s表示16PSK星座最小距离？

(b) 求α_M及β_M使16PSK在AWGN信道中的误码率可近似表示为

$$P_s \approx \alpha_M Q\left(\sqrt{\beta_M \gamma_s}\right)$$

(c) 利用（b）的结果求给定平均符号信噪比$\bar{\gamma}_s$时，16PSK在瑞利信道中的平均误码率的近似式。

(d) 假设采用格雷码，将（c）中得到的平均误码率转化为平均误比特率。

(e) 求在瑞利衰落信道中，达到误比特率为10^{-3}时BPSK和16PSK各自需要的$\bar{\gamma}_b$的近似值。从BPSK升到高阶调制16PSK时的功率损失是多大？

6.14 写出Nakagami-m衰落信道DPSK的平均误比特率的闭式表达式，并计算出$m=4$、$\bar{\gamma}_b = 10$dB时的平均误码率。

6.15 Nakagami衰落的参数m决定了衰落的特征，其取值范围从$m=0.5$到$m=\infty$。m表示直射信号功率和多径信号功率的比值，所以当$m=1$时，对应的信道为瑞利衰落信道，当$m=\infty$时，对应的就是无衰落的AWGN信道，$m=0.5$时的信道的性能比瑞利衰落还差。本题探

讨Nakagami衰落信道中，参数m对BPSK性能的影响。

请按信噪比范围0~20dB画出Nakagami信道中BPSK的平均误比特率（可用6.3.3节介绍的矩母函数的方法来得到平均误比特率）。参数m取1（瑞利）、2、4。当平均信噪比为10dB时，这些平均误比特率有什么区别？

6.16　某900MHz蜂窝系统中同时存在对数正态阴影衰落和瑞利衰落。调制方式为DPSK。运营商可以容忍0.01的中断率，即任意给定时间每100个客户中有1位不满意。非中断时，语音业务对平均误比特率的要求是$\bar{P}_b=10^{-3}$。假设噪声的功率谱密度为$N_0/2$，$N_0=10^{-16}$mW/Hz，信号的带宽为30kHz。假设路径损耗为自由空间传播模型，发送和接收均采用全向天线，阴影的标准差为$\sigma_{\psi_{dB}}=6$dB，移动终端的最大发射功率为100mW。求出小区的最大半径。

200

6.17　某蜂窝系统的小区是半径为100m的圆。假设传播条件满足为简化的路径损耗模型，其参数为$K=1$、$d_0=1$m、$\gamma=3$。假设存在瑞利衰落和标准差为$\sigma_{\psi_{dB}}=4$dB的对数正态阴影衰落。基站发射功率为$P_t=100$mW，系统的带宽为30kHz，噪声的功率谱密度为$N_0/2$，$N_0=10^{-14}$W/Hz，采用BPSK调制。我们的目的是想得到该小区的效覆率，即平均误比特率小于10^{-3}的位置占小区总面积的百分比。

(a) 求只经过路径损耗时，小区边界处的接收功率。

(b) 求使平均误比特率满足$\bar{P}_b<10^{-4}$的最小的平均接收功率（由路径损耗和阴影决定）。这里的平均误比特率是对瑞利衰落平均。

(c) 根据所给定的传播模型，求小区内满足$\bar{P}_b<10^{-4}$的位置面积占总面积的百分比。

6.18　本题推导DPSK在瑞利衰落信道中的误比特率。由对称性可知，发0和发1的错误率是相同的，故此我们假设在kT_b时刻发送0。这样，kT_b时刻发送的信号等于$(k-1)T_b$时刻的发送信号，即$s(k)=s(k-1)$。在快衰落信道中，对应的接收信号为：$r(k-1)=g_{k-1}s(k-1)+n(k-1)$和$r(k)=g_ks(k-1)+n(k)$，其中$g_{k-1}$和$g_k$分别是$(k-1)T_b$时刻和$kT_b$时刻对应的衰落信道的增益。

(a) 图5-20中相位比较器输入的判决量为

$$r(k)r^*(k-1)=g_kg_{k-1}^*+g_ks(k-1)^*n_{k-1}+g_{k-1}^*s_{k-1}^*n_k+n_kn_{k-1}^*$$

对于合理的信噪比，上式中最后一项$n_kn^*_{k-1}$可以忽略。定义$\tilde{n}_k=s_{k-1}^*n_k$、$\tilde{n}_{k-1}=s_{k-1}^*n_{k-1}$，于是我们得到一个新的随机变量$\tilde{z}=g_kg_{k-1}^*+g_k\tilde{n}_{k-1}^*+g_{k-1}^*\tilde{n}_k$。给定$kT_b$时刻发送0时，若$x=\text{Re}\{\tilde{z}\}<0$就会发生判决错误。因此我们必须要得到$x$的分布。$x$的特征函数是其概率密度函数的双边拉普拉斯变换：

$$\Phi_X(s)=\int_{-\infty}^{\infty}p_X(s)e^{-sx}\,dx=E[e^{-sx}]$$

上式有一个左半平面极点p_1和一个右半平面极点p_2，故可写成

$$\Phi_X(s)=\frac{p_1p_2}{(s-p_1)(s-p_2)}$$

左半平面极点p_1对应$p_X(x)$，$x\geqslant0$，右半平面极点p_2对应$p_X(x)$，$x<0$。

(b) 用部分分式展开的方法将$\Phi_X(s)$写成

$$\Phi_X(s)=\frac{p_1p_2}{p_1-p_2}\frac{1}{s-p_1}+\frac{p_1p_2}{p_2-p_1}\frac{1}{s-p_2}$$

判决错误发生在$x=\mathrm{Re}\{\tilde{z}\}<0$时，故我们只需考虑$\Phi_X(s)$的第二项对应的概率密度函数$p_X(x)$，$x<0$。

201

（c）对（b）中$\Phi_x(s)$的第二项进行反拉普拉斯变换可得

$$p_X(x)=\frac{p_1p_2}{p_2-p_1}e^{p_2x} \quad x<0$$

（d）由（c）可证$P_b=-p_1/(p_2-p_1)$。

$x=\mathrm{Re}\{\tilde{z}\}=\mathrm{Re}\{g_kg_{k-1}^*+g_k\tilde{n}_{k-1}^*+g_{k-1}^*\tilde{n}_k\}$中的信道增益$g_k$和$g_{k-1}$，以及噪声$\tilde{n}_k$和$\tilde{n}_{k-1}$都是复高斯随机变量。所以可用复高斯随机变量的广义二次型解出极点p_1和p_2（参见文献[1]的附录B及文献[12]的附录B）：

$$p_1=\frac{-1}{N_0(\bar{\gamma}_b[1+\rho_c]+1)} \qquad p_2=\frac{1}{N_0(\bar{\gamma}_b[1-\rho_c]+1)}$$

式中的ρ_C是经过T_b时刻的信道相关系数。

（e）用（d）中得到的p_1和p_2，求出瑞利快衰落信道中误比特率P_b的一般表达式。

（f）从（e）的结果的简化得到慢衰落信道中的平均误比特率$\bar{P}_b=1/[2(1+\bar{\gamma}_b)]$。

6.19 按$B_DT=0.01$、0.001、0.0001分别画出瑞利快衰落信道中DPSK的误比特率曲线。$\rho_C=J_0(2\pi B_DT)$，$\bar{\gamma}_b$取值范围0~60dB。对每一种B_DT的取值，$\bar{\gamma}_b$的值大约为多少时背景误码开始起主导作用？

6.20 假设高斯多普勒功率谱、$B_D=80\text{Hz}$、莱斯因子$K=2$、数据速率40kbit/s，求DQPSK因多普勒频移而引起的背景误码率。

6.21 某无线信道的平均时延扩展为100ns，多普勒扩展为80Hz。假设均匀散射的瑞利衰落，调制方式为DQPSK。考虑由多普勒频移和ISI引起的背景误码，数据速率大致在什么范围内时，可保证误比特率小于10^{-4}？

6.22 假设时延扩展为$\sigma_{T_m}=3\mu s$，要求的误比特率P_b小于10^{-3}，请利用图6-5求出BPSK、QPSK或MSK最大可传输的数据速率。

参考文献

[1] J. G. Proakis, *Digital Communications*, 4th ed., McGraw-Hill, New York, 2001.
[2] M. K. Simon, S. M. Hinedi, and W. C. Lindsey, *Digital Communication Techniques: Signal Design and Detection*, Prentice-Hall, Englewood Cliffs, NJ, 1995.
[3] S. Rhodes, "Effect of noisy phase reference on coherent detection of offset-QPSK signals," *IEEE Trans. Commun.*, pp. 1046–55, August 1974.
[4] N. R. Sollenberger and J. C.-I. Chuang, "Low-overhead symbol timing and carrier recovery for portable TDMA radio systems," *IEEE Trans. Commun.*, pp. 1886–92, October 1990.
[5] R. Pawula, S. Rice, and J. Roberts, "Distribution of the phase angle between two vectors perturbed by Gaussian noise," *IEEE Trans. Commun.*, pp. 1828–41, August 1982.
[6] W. Cowley and L. Sabel, "The performance of two symbol timing recovery algorithms for PSK demodulators," *IEEE Trans. Commun.*, pp. 2345–55, June 1994.
[7] W. T. Webb and L. Hanzo, *Modern Quadrature Amplitude Modulation*, IEEE/Pentech Press, London, 1994.
202
[8] X. Tang, M.-S. Alouini, and A. Goldsmith, "Effect of channel estimation error on M-QAM BER

performance in Rayleigh fading," *IEEE Trans. Commun.*, pp. 1856–64, December 1999.

[9] M. K. Simon and M.-S. Alouini, *Digital Communication over Fading Channels: A Unified Approach to Performance Analysis,* Wiley, New York, 2000.

[10] S. Hinedi, M. Simon, and D. Raphaeli, "The performance of noncoherent orthogonal M-FSK in the presence of timing and frequency errors," *IEEE Trans. Commun.*, pp. 922–33, February–April 1995.

[11] E. Grayver and B. Daneshrad, "A low-power all-digital FSK receiver for deep space applications," *IEEE Trans. Commun.*, pp. 911–21, May 2001.

[12] M. Schwartz, W. R. Bennett, and S. Stein, *Communication Systems and Techniques,* McGraw-Hill, New York, 1966 [reprinted 1995 by Wiley/IEEE Press].

[13] J. Craig, "New, simple and exact result for calculating the probability of error for two-dimensional signal constellations," *Proc. Military Commun. Conf.*, pp. 25.5.1–25.5.5, November 1991.

[14] F. S. Weinstein, "Simplified relationships for the probability distribution of the phase of a sine wave in narrow-band normal noise," *IEEE Trans. Inform. Theory,* pp. 658–61, September 1974.

[15] M. K. Simon and D. Divsalar, "Some new twists to problems involving the Gaussian probability integral," *IEEE Trans. Commun.*, pp. 200–10, February 1998.

[16] R. F. Pawula, "A new formula for MDPSK symbol error probability," *IEEE Commun. Lett.*, pp. 271–2, October 1998.

[17] P. A. Bello and B. D. Nelin, "The influence of fading spectrum on the bit error probabilities of incoherent and differentially coherent matched filter receivers," *IEEE Trans. Commun. Syst.*, pp. 160–8, June 1962.

[18] P. Y. Kam, "Tight bounds on the bit-error probabilities of 2DPSK and 4DPSK in nonselective Rician fading," *IEEE Trans. Commun.*, pp. 860–2, July 1998.

[19] P. Y. Kam, "Bit error probabilities of MDPSK over the nonselective Rayleigh fading channel with diversity reception," *IEEE Trans. Commun.*, pp. 220–4, February 1991.

[20] V. Fung, R. S. Rappaport, and B. Thoma, "Bit error simulation for $\pi/4$ DQPSK mobile radio communication using two-ray and measurement based impulse response models," *IEEE J. Sel. Areas Commun.*, pp. 393–405, April 1993.

[21] S. Chennakeshu and G. J. Saulnier, "Differential detection of $\pi/4$-shifted-DQPSK for digital cellular radio," *IEEE Trans. Veh. Tech.*, pp. 46–57, February 1993.

[22] P. A. Bello and B. D. Nelin, "The effects of frequency selective fading on the binary error probabilities of incoherent and differentially coherent matched filter receivers," *IEEE Trans. Commun. Syst.*, pp. 170–86, June 1963.

[23] M. B. Pursley, *Introduction to Digital Communications,* Prentice-Hall, Englewood Cliffs, NJ, 2005.

[24] S. Gurunathan and K. Feher, "Multipath simulation models for mobile radio channels," *Proc. IEEE Veh. Tech. Conf.*, pp. 131–4, May 1992.

[25] J. C.-I. Chuang, "The effects of time delay spread on portable radio communications channels with digital modulation," *IEEE J. Sel. Areas Commun.*, pp. 879–89, June 1987.

[26] C. Liu and K. Feher, "Bit error rate performance of $\pi/4$ DQPSK in a frequency selective fast Rayleigh fading channel," *IEEE Trans. Veh. Tech.*, pp. 558–68, August 1991.

203

第7章　分　　集

从第6章中我们知道，瑞利衰落和对数正态阴影衰落都会使调制性能产生很大的功率损失。减轻衰落影响的有效技术之一就是对独立的衰落信号进行分集合并。分集（diversity）就是在独立的衰落路径上发送相同的数据，由于独立路径在同一时刻经历深衰落的概率很小，因此经过适当的合并后，接收信号的衰落程度就会被减小。例如一个装有两根发送或接收天线的系统，如果天线间距足够远，那么两根天线同时经历深衰落的可能性很小。如果我们选择信号最强的那个天线，就能获得比单天线时更好的信号，此即选择合并（selection combining）技术。本章讨论常用的发送和接收分集方法。文献[1]的9.10节还介绍了其他一些在性能和复杂性方面都有所长的分集技术。

用来对抗多径衰落的分集技术叫做微分集（microdiversity），这是本章讨论的重点。用来对抗楼房等物体的阴影效应的分集叫做宏分集（macrodiversity）。宏分集一般是将几个基站或接入点的接收信号进行合并，这样做需要不同的基站或接入点进行协调。对于有基础设施的架构式无线网络，这样的协调是网络协议的一部分。我们将把宏分集的内容推后到第15章，在那里我们将详细讨论这种网络的设计。

7.1　独立衰落路径的实现

在无线通信系统中有很多方法可以实现独立的衰落路径。其中之一是使用多个发送天线或接收天线，也就是使用天线阵列，其阵元之间有一定的空间距离。这种形式的分集叫做空间分集（space diversity）。在接收空间分集中，实现独立的衰落路径不需要增加发送功率或带宽，通过分集信号进行相干合并还能提高接收信噪比，这个相对于单天线的信噪比增益叫做阵列增益。通过对天线进行适当的加权，发送分集也能获得阵列增益。除阵列增益外，空间分集还能带来分集增益。分集增益定义为分集合并后误码率斜率的变化。后面几节将针对具体的分集合并技术讨论其阵列增益和分集增益。

204

无论是发送空间分集还是接收空间分集，为了获得最大的分集增益，一般要求有足够的天线间距以使各天线上的衰落近似独立。由式（3-26）可知，对于均匀散射环境及全向的发送和接收天线，达到衰落独立需要的最小间距近似为波长的一半（精确值是波长的0.38倍）。如果发送或接收天线是定向的（这在扇区化基站中是很常见的），多径成分将主要集中在直射线周围的一个小角度内，因此需要更大的天线间距才能获得独立的衰落[2]。

另一种方法是使用不同极化方式（如水平极化和垂直极化）的两根发送或接收天线，不同极化方式的两路信号经由相同的路径传播，传播环境中的许多随机反射将把信号功率大致均匀地分配在两个极化方向上，于是按两种不同极化接收的功率近似相同。因为散射角相对于每个极化方向是任意的，所以不同极化的接收功率同时经历深衰落的可能性很小。极化分集有两个缺点：首先，对应于两种极化方向，最多只能有两个分集支路。其次，因为发送或接收功率要分配到两个极化天线上，所以极化分集有3dB的功率损失。

号等于原始发送的信号$s(t)$乘以一个随机的复系数$\alpha_\Sigma=\sum_i a_i r$。合并输出的信噪比$\gamma_\Sigma$也是一个随机变量，其概率分布取决于分集路径数、各路径上的衰落分布以及合并方式。 206

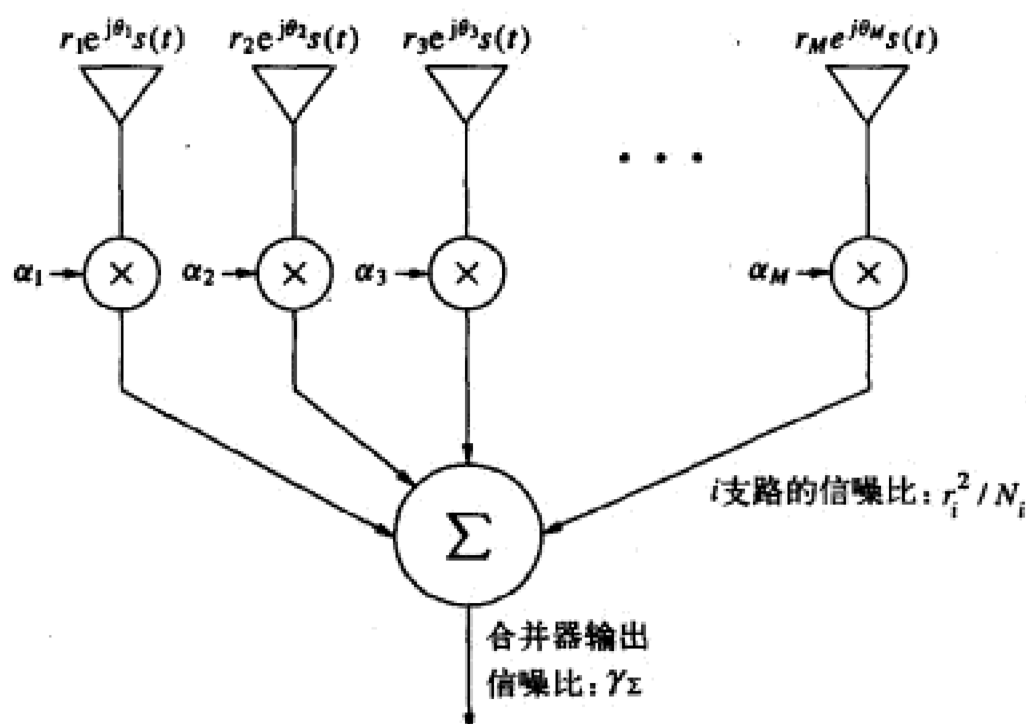


图7-1 线性合并器

对多路接收信号进行相干合并带来了接收空间分集的阵列增益。即使没有衰落，相干合并也能增加平均接收信噪比。例如，假设没有衰落，$r_i=\sqrt{E_s}$（E_s是发送信号的符号能量）。假设每个支路上的噪声功率谱密度都是$N_0/2$，脉冲成形满足$BT_s=1$，则每个支路都有相同的信噪比$\gamma_i=E_s/N_0$，令$a_i=r_i/\sqrt{N_0}$（后面将看到这是最大比合并的最佳加权系数），那么合并后的接收信噪比为

$$\gamma_\Sigma=\frac{\left(\sum_{i=1}^{M}a_ir_i\right)^2}{N_0\sum_{i=1}^{M}a_i^2}=\frac{\left(\sum_{i=1}^{M}\frac{E_s}{\sqrt{N_0}}\right)^2}{N_0\sum_{i=1}^{M}\frac{E_s}{N_0}}=\frac{ME_s}{N_0} \tag{7-1}$$

可见没有衰落时，通过适当的加权对来自不同天线的M路接收信号进行相干合并后，可以使信噪比增大M倍。无衰落时的这种信噪比增益称作**阵列增益**（array gain）。其确切定义是指合并输出的平均信噪比$\bar{\gamma}_\Sigma$相对于支路的平均信噪比$\bar{\gamma}$的增益：

$$A_g=\frac{\bar{\gamma}_\Sigma}{\bar{\gamma}}$$

所有的分集合并方式都有阵列增益，最大比合并的阵列增益最大。由于阵列增益的原因，在相同的平均信噪比下，采用了多个发送或接收天线的分集系统在衰落信道下的性能比无分集的系统在AWGN信道下的性能好。我们将在后面给出的最大比合并和等增益合并的性能曲线中看到这一点。 207

在衰落信道中，多个衰落路径合并后会有一个更好的概率分布。无论是空间分集还

号等于原始发送的信号$s(t)$乘以一个随机的复系数$\alpha_\Sigma=\sum_i a_i r$。合并输出的信噪比$\gamma_\Sigma$也是一个随机变量，其概率分布取决于分集路径数、各路径上的衰落分布以及合并方式。 206

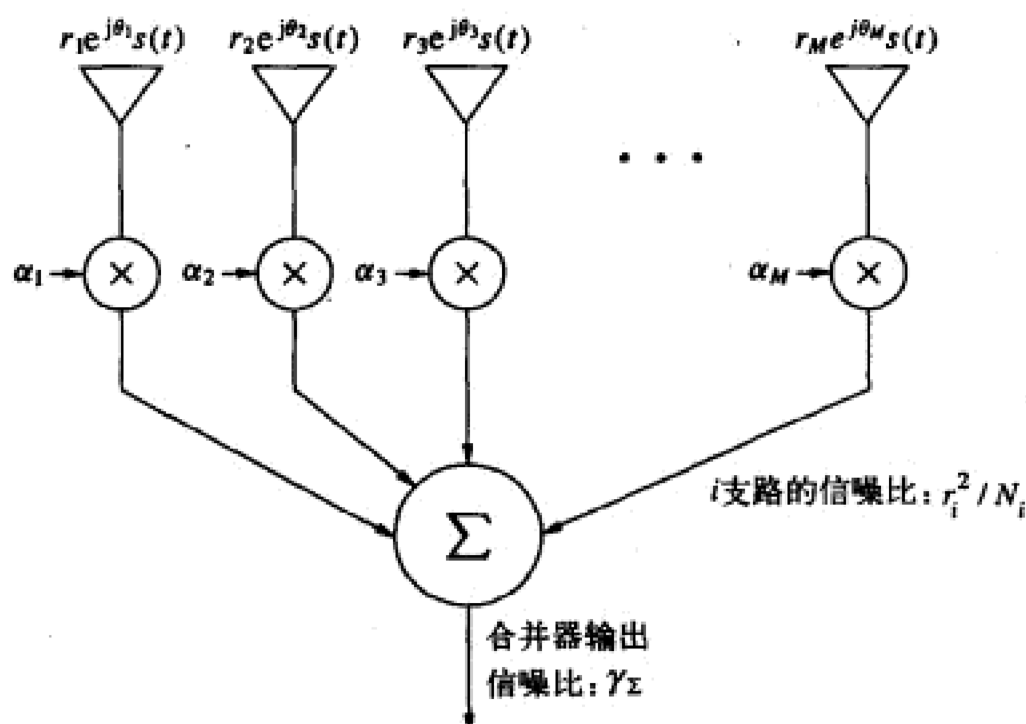


图7-1 线性合并器

对多路接收信号进行相干合并带来了接收空间分集的阵列增益。即使没有衰落，相干合并也能增加平均接收信噪比。例如，假设没有衰落，$r_i=\sqrt{E_s}$（E_s是发送信号的符号能量）。假设每个支路上的噪声功率谱密度都是$N_0/2$，脉冲成形满足$BT_s=1$，则每个支路都有相同的信噪比$\gamma_i=E_s/N_0$，令$a_i=r_i/\sqrt{N_0}$（后面将看到这是最大比合并的最佳加权系数），那么合并后的接收信噪比为

$$\gamma_\Sigma=\frac{\left(\sum_{i=1}^{M}a_i r_i\right)^2}{N_0\sum_{i=1}^{M}a_i^2}=\frac{\left(\sum_{i=1}^{M}\frac{E_s}{\sqrt{N_0}}\right)^2}{N_0\sum_{i=1}^{M}\frac{E_s}{N_0}}=\frac{ME_s}{N_0} \tag{7-1}$$

可见没有衰落时，通过适当的加权对来自不同天线的M路接收信号进行相干合并后，可以使信噪比增大M倍。无衰落时的这种信噪比增益称作*阵列增益*（array gain）。其确切定义是指合并输出的平均信噪比$\bar{\gamma}_\Sigma$相对于支路的平均信噪比$\bar{\gamma}$的增益：

$$A_g=\frac{\bar{\gamma}_\Sigma}{\bar{\gamma}}$$

所有的分集合并方式都有阵列增益，最大比合并的阵列增益最大。由于阵列增益的原因，在相同的平均信噪比下，采用了多个发送或接收天线的分集系统在衰落信道下的性能比无分集的系统在AWGN信道下的性能好。我们将在后面给出的最大比合并和等增益合并的性能曲线中看到这一点。 207

在衰落信道中，多个衰落路径合并后会有一个更好的概率分布。无论是空间分集还

是其他方式，分集系统在衰落信道中的性能指标$\bar{P}_s$和P_{out}可按6.3.1节和6.3.2节中的方式定义为

$$\bar{P}_s = \int_0^\infty P_s(\gamma) p_{\gamma_\Sigma}(\gamma)\,d\gamma \tag{7-2}$$

$$P_{out} = p(\gamma_\Sigma \leqslant \gamma_0) = \int_0^{\gamma_0} p_{\gamma_\Sigma}(\gamma)\,d\gamma \tag{7-3}$$

式中的$P_s(\gamma)$是信噪比为γ的AWGN信道的误码率，γ_0是目标信噪比。多径合并能改善信噪比的分布，从而降低$\bar{P}_s$和P_{out}，由此带来的性能增益叫做*分集增益*（diversity gain）。特别是某些分集系统的误码率可以表示成$\bar{P}_s = c\bar{\gamma}^{-M}$的形式，其中$c$是取决于具体调制编码方式的常数，$\bar{\gamma}$是每个支路的平均接收信噪比，$M$称为系统的*分集阶数*（diversity order）。分集阶数是分集系统的平均误码率随平均信噪比变化的曲线的斜率（见后面的图7-3和图7-6）。对比式（6-61），无分集瑞利衰落信道的平均误码率近似为$\bar{P}_s \approx \alpha_M/(2\beta_M\bar{\gamma})$，其分集阶数为1，对应于一根接收天线。有$M$根天线的分集系统的最大分集阶数是$M$，若分集阶数实际等于$M$，则称此系统达到了*满分集阶数*（full diversity order）。

下面将详细介绍不同的合并方式及其性能。这些合并方式体现了不同的性能和复杂度的折中。

7.2.2　选择合并

选择合并（selection combining，SC）输出信噪比最高的那个支路上的信号。若各支路的噪声功率都相等，即$N_i=N$，则等价于选择$r_i^2+N_i$最大的支路[1]。因为每一个时刻只用一个支路，选择合并只需要一个接收机，随时切换到被选的天线支路上即可。不过对于连续发送的系统，可能需要每个天线支路都安装一个接收机来监测各支路上的信噪比。选择合并输出的信噪比等于各支路信噪比的最大值。此外，因为不需要各支路同相，故此SC可以采用相干调制或差分调制。

对于M支路的分集，γ_Σ的累计分布函数为：

208

$$P_{\gamma_\Sigma}(\gamma) = p(\gamma_\Sigma < \gamma) = p(\max[\gamma_1, \gamma_2, \cdots, \gamma_M] < \gamma) = \prod_{i=1}^{M} p(\gamma_i < \gamma) \tag{7-4}$$

对$P_{\gamma_\Sigma}(\gamma)$中的$\gamma$进行微分可得到$\gamma_\Sigma$的概率密度函数。将$\gamma=\gamma_0$代入$P_{\gamma_\Sigma}(\gamma)$可得到中断率。假设$M$个不相关瑞利衰落的幅度是$r_i$，则$i$支路的瞬时信噪比为$\gamma_i = r_i^2/N$。$\gamma_i$服从指数分布：

$$p(\gamma_i) = \frac{1}{\bar{\gamma}_i} e^{-\gamma_i/\bar{\gamma}_i} \tag{7-5}$$

其中$\bar{\gamma}_i$是i支路的平均信噪比：$\bar{\gamma}_i = E[\gamma_i]$。由式（6-47）可得，瑞利衰落下$i$支路对应于目标$\gamma_0$的中断率为：

$$P_{out}(\gamma_0) = 1 - e^{-\gamma_0/\bar{\gamma}_i} \tag{7-6}$$

1. 实际中，$r_i^2+N_i$比信噪比更容易测量，因为前者只是测量接收到的总功率。

对应目标γ_0，SC的中断率为

$$P_{out}(\gamma_0) = \prod_{i=1}^{M} p(\gamma_i < \gamma_0) = \prod_{i=1}^{M} [1 - e^{-\gamma_0/\bar{\gamma}_i}] \tag{7-7}$$

如果所有支路的平均信噪比相同（对所有i，$\bar{\gamma}_i = \bar{\gamma}$），上式可简化为

$$P_{out}(\gamma_0) = p(\gamma_\Sigma < \gamma_0) = [1 - e^{-\gamma_0/\bar{\gamma}}]^M \tag{7-8}$$

对式（7-8）中的γ_0微分得到γ_Σ的概率密度函数为

$$p_{\gamma_\Sigma}(\gamma) = \frac{M}{\bar{\gamma}}[1 - e^{-\gamma/\bar{\gamma}}]^{M-1} e^{-\gamma/\bar{\gamma}} \tag{7-9}$$

从式（7-9）可得到独立同分布瑞利衰落下，合并输出的平均信噪比为

$$\begin{aligned} \bar{\gamma}_\Sigma &= \int_0^\infty \gamma p_{\gamma_\Sigma}(\gamma) d\gamma \\ &= \int_0^\infty \frac{\gamma M}{\bar{\gamma}}[1 - e^{-\gamma/\bar{\gamma}}]^{M-1} e^{-\gamma/\bar{\gamma}} d\gamma \\ &= \bar{\gamma} \sum_{i=1}^{M} \frac{1}{i} \end{aligned} \tag{7-10}$$

209

因此，平均信噪比增益以及相应的阵列增益随M的增加而增加。但这种增加不是线性增加，从无分集到两支路分集得到的增益最多。分集支路数从二条增加到三条所获得的增益大大小于从一条增加到二条得到的增益。一般来说，随着M进一步增大，阵列增益的增加值逐步趋于零。图7-2示出了这种趋势。图中画出了独立同分布瑞利衰落下不同M时

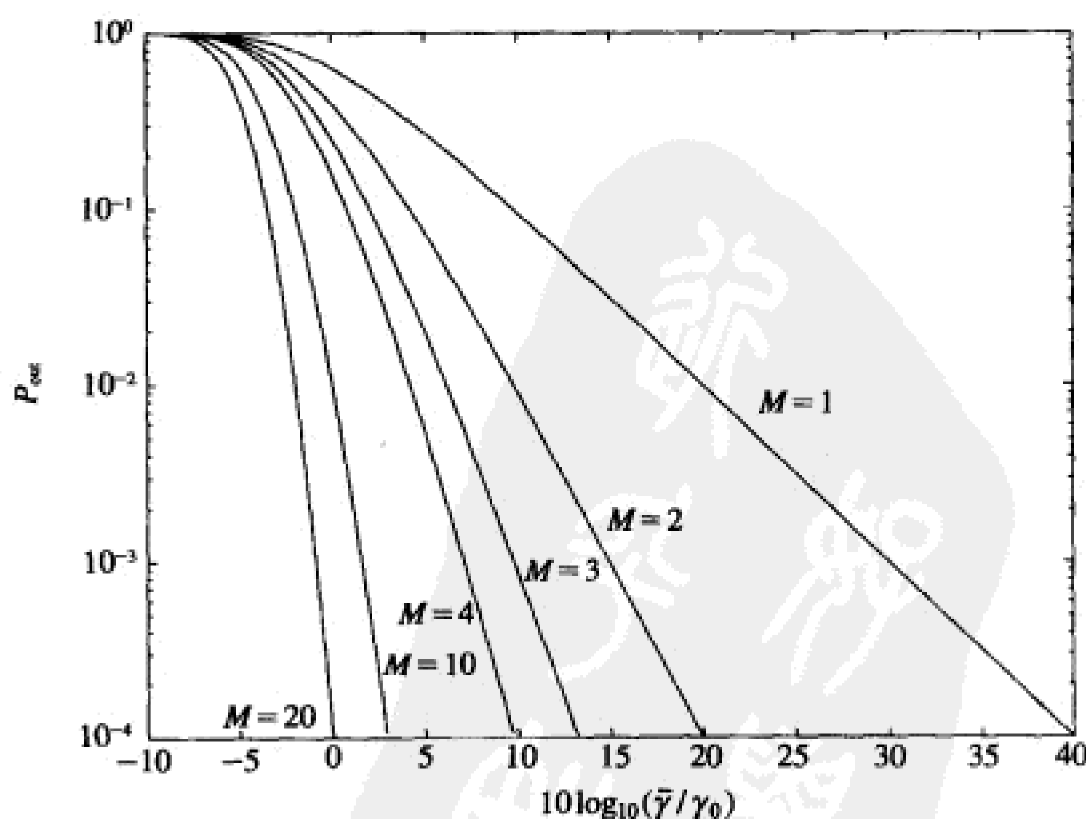


图7-2 瑞利衰落下选择合并的中断率

P_{out}和 $\bar{\gamma}/\gamma_0$ 的关系曲线。可以看到当支路从一个增加到两个时，SC的性能有很大的提高，$M=1$增加到$M=2$时，对应1%中断率所需的信噪比大概降低了12dB，0.01%中断率时降低了20dB。但从两支路增加到三支路时，对应0.01%的中断率所需的平均信噪比只降低了7dB左右。从三支路增加到四支路时只降低了4dB。很明显，从无分集到两支路分集节约的功率最多，随着支路数逐步增加，额外所得的增益逐步减小并趋于消失。另外需要注意的一点是，即便各支路都是瑞利衰落，合并输出的信噪比不再服从指数分布。

例7.1：设信道为瑞利衰落，调制方式为BPSK调制，可接受的误码率为$P_b=10^{-3}$。求$M=1$（无分集）、$M=2$及$M=3$时选择分集的中断率。假设各支路的平均信噪比相同，为$\bar{\gamma}=15$dB。

解：BPSK调制信号在$\gamma_0=7$dB时达到$p_b=10^{-3}$，因此$\gamma_0=7$dB，将$\gamma_0=10^{0.7}$和$\bar{\gamma}=10^{1.5}$带入式（7-8）得到，$M=1$时$P_{out}=0.1466$，$M=2$时$P_{out}=0.0215$，$M=3$时$P_{out}=0.0031$。可以看到，每增加一个支路时，中断率大约降低一个数量级。

平均误码率可用合成后的信噪比分布 $p_{\gamma_\Sigma}(\gamma)$ 由式（7-2）得到，其中$P_s(\gamma)$是AWGN信道的误码率。相干调制在大多数衰落分布下没有闭式解，只能用数值方法或近似方法得到结果。图7-3画出了独立同分布瑞利衰落下 $\bar{P}_b$ 和 $\bar{\gamma}_b$ 的关系，这个曲线是对 $\int Q(\sqrt{2\gamma})p_{\gamma_\Sigma}(\gamma)d\gamma$ 进行数值积分得到的，其中的 $p_{\gamma_\Sigma}(\gamma)$ 由式（7-9）给出。从图中可见，$M\geqslant 8$的分集系统的误码率比相同信噪比的AWGN信道还要低，这正是因为分集有阵列增益，最大比合并和等增益合并也如此。当每个支路都是独立同分布的瑞利衰落时，差分调制的平均误码率存在闭式解，见文献[1]的9.7节及文献[4]的6.1节。例如DPSK在式（7-9）中给出的 $p_{\gamma_\Sigma}(\gamma)$ 分布下的平均误码率为

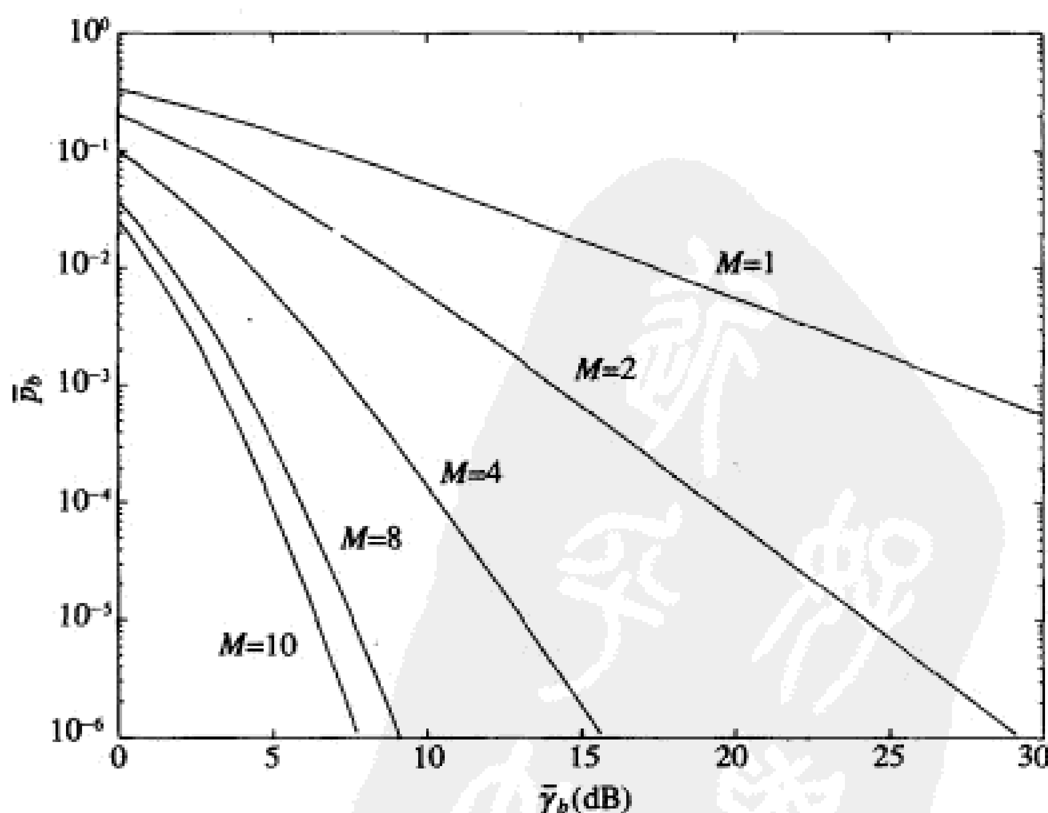


图7-3　独立同分布的瑞利衰落下选择合并时BPSK的平均P_b

$$\bar{P}_b=\int_0^{\infty}\frac{1}{2}\mathrm{e}^{-\gamma}p_{\gamma_\Sigma}(\gamma)\,\mathrm{d}\gamma=\frac{M}{2}\sum_{m=0}^{M-1}(-1)^m\frac{\binom{M-1}{m}}{1+m+\bar{\gamma}} \tag{7-11}$$

210

前面的推导中假设各支路的幅度不相关。相关性会造成轻微的性能损失，这种损失在相关系数低于0.5时可以忽略不计。文献[1]的9.7节及文献[2]推导了支路相关性带来的性能损失的精确值。

7.2.3 门限合并

对于连续发送的系统，选择合并需要每个支路都有一个接收机来连续监测支路的信噪比。门限合并是一种更简单的合并方法，它用同一个接收机顺序监测每个支路，输出第一个信噪比高于门限值γ_T的信号，从而避免了在每个支路上都安装一个接收机。门限合并也和选择合并一样，每个时刻只有一路信号输出，不需要各路同相，因此相干调制和差分调制都可以采用门限合并。

一旦选定支路后，只要该支路的信噪比一直高于门限值，合并器就始终输出这个支路上的信号。当这一支路的信噪比低于门限值时，就切换到其他支路。决定切换到哪一个其他支路上有一些不同的方法[5]，最简单就是随机切换。只有两个支路时，就是直接切换到另一个支路上。这种方法叫做*切停合并*（switch-and-stay combining，SSC）。图7-4示出了切换的过程及相应的信噪比。SSC并不是选择信噪比最高的支路，因此它的性能介于无分集和理想选择合并之间。

211

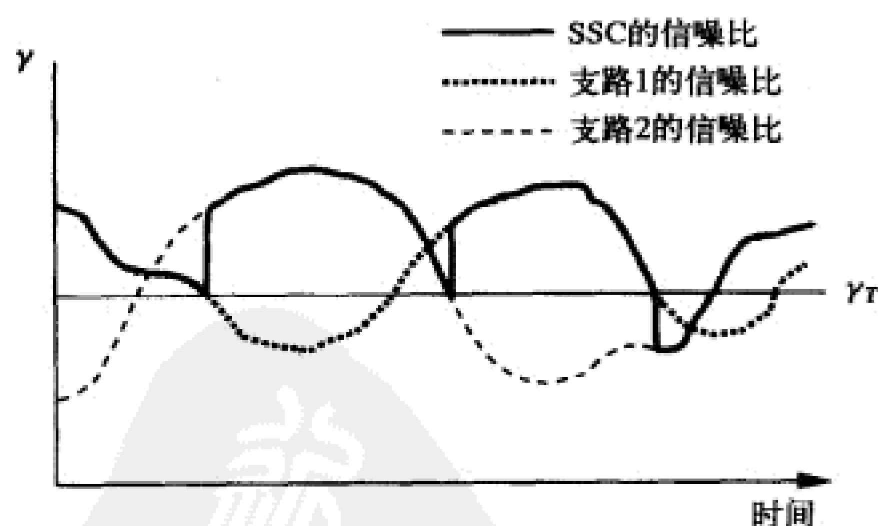


图7-4 SSC中的信噪比

记i支路的信噪比为γ_i，合并输出的信噪比为γ_Σ。γ_Σ的累积分布函数取决于门限值γ_T和γ_i的累积分布函数。对于独立同分布的两支路分集，合并输出的累积分布函数$P_{\gamma_\Sigma}(\gamma)=p(\gamma_\Sigma\leqslant\gamma)$可以用支路信噪比的累积分布函数$P_{\gamma_i}(\gamma)=p(\gamma_i\leqslant\gamma)$和概率密度函数$p_{\gamma_i}(\gamma)$表示为

$$P_{\gamma_\Sigma}(\gamma)=\begin{cases}P_{\gamma_1}(\gamma_T)P_{\gamma_2}(\gamma) & \gamma<\gamma_T\\ p(\gamma_T\leqslant\gamma_2\leqslant\gamma)+P_{\gamma_1}(\gamma_T)P_{\gamma_2}(\gamma) & \gamma\geqslant\gamma_T\end{cases} \tag{7-12}$$

若每个支路都是同分布的瑞利衰落，$\bar{\gamma}_i=\bar{\gamma}(i=1,2)$，则有

$$P_{\gamma_\Sigma}(\gamma)=\begin{cases}1-\mathrm{e}^{-\gamma_T/\bar{\gamma}}-\mathrm{e}^{-\gamma/\bar{\gamma}}+\mathrm{e}^{-(\gamma_T+\gamma)/\bar{\gamma}} & \gamma<\gamma_T\\ 1-2\mathrm{e}^{-\gamma/\bar{\gamma}}+\mathrm{e}^{-(\gamma_T+\gamma)/\bar{\gamma}} & \gamma\geqslant\gamma_T\end{cases} \tag{7-13}$$

对应目标γ_0的中断率就是将$\gamma=\gamma_0$代入$P_{\gamma_\Sigma}(\gamma)$：

$$P_{\text{out}}(\gamma_0)=P_{\gamma_\Sigma}(\gamma_0)=\begin{cases}1-\mathrm{e}^{-\gamma_T/\bar{\gamma}}-\mathrm{e}^{-\gamma_0/\bar{\gamma}}+\mathrm{e}^{-(\gamma_T+\gamma_0)/\bar{\gamma}} & \gamma_0<\gamma_T\\ 1-2\mathrm{e}^{-\gamma_0/\bar{\gamma}}+\mathrm{e}^{-(\gamma_T+\gamma_0)/\bar{\gamma}} & \gamma_0\geqslant\gamma_T\end{cases} \tag{7-14}$$

关于SSC在其他衰落类型下的性能，以及衰落相关性的影响，请参考文献[1]的9.8节及文献[6~7]。文献[1]的9.8节指出，对任意的衰落分布，具有最佳门限$\gamma_T=\gamma_0$的SSC的中断率与SSC相同。

212

例7.2：考虑两支路的SSC，每个支路都是独立同分布的瑞利衰落，支路的平均信噪比为$\bar{\gamma}=15$dB。假设使用BPSK调制，可接受误码率为$P_b=10^{-3}$。求门限值分别为$\gamma_T=5$dB、7dB和10dB时的中断率。讨论γ_T的值对中断率的影响。结合例7.1比较SSC、SC及无分集时的中断率。

解：同例7.1，$\gamma_0=7$dB。$\gamma_T=5$dB时，由$\gamma_0\geqslant\gamma_T$可根据式（7-14）的第二行得到

$$P_{\text{out}}=1-2\mathrm{e}^{-10^{0.7}/10^{1.5}}+\mathrm{e}^{-(10^{0.5}+10^{0.7})/10^{1.5}}=0.0654$$

$\gamma_T=7$dB时，$\gamma_0=\gamma_T$，同样可根据式（7-14）的第二行得到

$$P_{\text{out}}=1-2\mathrm{e}^{-10^{0.7}/10^{1.5}}+\mathrm{e}^{-(10^{0.7}+10^{0.7})/10^{1.5}}=0.0215$$

$\gamma_T=10$dB时，$\gamma_0<\gamma_T$，可从式（7-14）的第一行得到

$$P_{\text{out}}=1-\mathrm{e}^{-10/10^{1.5}}-\mathrm{e}^{-10^{0.7}/10^{1.5}}+\mathrm{e}^{-(10+10^{0.7})/10^{1.5}}=0.0397$$

可见，当$\gamma_T=7$dB时的中断率小于其他两种情况。$\gamma_T=5$dB时，门限值偏低，当前选中的支路有可能长时间低于目标值γ_0而未被切走，致使中断率提高。$\gamma_T=10$dB时，门限值偏高，当前输出的支路经常低于门限值，这会使其立即被切换到另一路，而另一路的信噪比可能更低。本例说明能使P_{out}最小的门限值γ_T就是目标值γ_0。

根据例7.1中的数据，SC的$P_{\text{out}}=0.0215$，因此$\gamma_T=7$dB是SSC的最佳门限，此时的性能和SC一样。门限值偏离最佳值时的性能要比SC差很多。但这三种门限值的SSC性能都比例7.1中无分集的$P_{\text{out}}=0.1466$好。

对式（7-12）中的γ进行微分可得到SSC输出信噪比γ_Σ的概率密度函数$p_{\gamma_\Sigma}(\gamma)$，然后可用式（7-2）得到平均误码率，其中$P_s(\gamma)$是AWGN信道下的误码率。在大多数衰落分布下，相干解调没有闭式结果，只能通过数值计算或近似计算求得。对于独立同分布的瑞利衰落，对式（7-13）微分得到

$$p_{\gamma_\Sigma}(\gamma)=\begin{cases}(1-\mathrm{e}^{-\gamma_T/\bar{\gamma}})(1/\bar{\gamma})\mathrm{e}^{-\gamma/\bar{\gamma}} & \gamma<\gamma_T\\ (2-\mathrm{e}^{-\gamma_T/\bar{\gamma}})(1/\bar{\gamma})\mathrm{e}^{-\gamma/\bar{\gamma}} & \gamma\geqslant\gamma_T\end{cases} \tag{7-15}$$

和SC一样，采用SSC的相干解调在大多数衰落分布下，平均误码率没有闭式解，只能进行数值计算。但是当每个支路都是独立同分布的瑞利衰落时，差分调制的误码率存在闭式解。DPSK的平均误码率为

213

$$\bar{P}_b=\int_0^\infty\frac{1}{2}\mathrm{e}^{-\gamma}p_{\gamma_\Sigma}(\gamma)\,\mathrm{d}\gamma=\frac{1}{2(1+\bar{\gamma})}(1-\mathrm{e}^{-\gamma_T/\bar{\gamma}}+\mathrm{e}^{-\gamma_T}\mathrm{e}^{-\gamma_T/\bar{\gamma}}) \tag{7-16}$$

例7.3：考虑两支路的SSC，每个支路都是独立同分布的瑞利衰落，支路的平均信噪比为 $\bar{\gamma}=15\text{dB}$。求门限值分别为 $\gamma_T=3\text{dB}$、7dB和10dB时DPSK的平均误码率。讨论不同 γ_T 下对误码率的影响。比较SSC、SC及无分集时的误码率。

解：在式（7-16）中代入 $\bar{\gamma}=15\text{dB}$ 及不同的门限 $\gamma_T=3\text{dB}$、7dB、10dB，分别得到 $\bar{P}_b=0.0029$、$\bar{P}_b=0.0023$ 及 $\bar{P}_b=0.0042$。如同前例，最佳门限能使平均误码率最小，门限过高或过低都会使性能下降。从式（7-11）可得SC的平均误码率是 $\bar{P}_b=0.5(1+10^{1.5})^{-1}-0.5(2+10^{1.5})^{-1}=4.56\times10^{-4}$，大约比最佳门限的SSC低一个数量级。无分集时的平均误码率是 $\bar{P}_b=0.5(1+10^{1.5})^{-1}=0.0153$，比两支路的SSC的高约一个数量级。

7.2.4 最大比合并

SC和SSC输出某一个支路上的信号，而最大比合并（maximal-ratio combining，MRC）输出的则是各支路信号的加权和，即图7-1中的 α_i 都不为零。各个支路同相相加，因此 $\alpha_i=a_i e^{-j\theta_i}$，其中 θ_i 等于 i 支路信号的相位。合并输出的包络是 $r=\sum_{i=1}^{M}a_i r_i$。假设每个支路的噪声功率谱密度都是 $N_0/2$，则合并输出的总噪声功率谱密度是 $N_{\text{tot}}/2=\sum_{i=1}^{M}a_i N_0/2$。这样，可得合并输出的信噪比为：

$$\gamma_\Sigma=\frac{r^2}{N_{\text{tot}}}=\frac{1}{N_0}\frac{\left(\sum_{i=1}^{M}a_i r_i\right)^2}{\sum_{i=1}^{M}a_i^2} \tag{7-17}$$

我们的目标是选择合适的 a_i 使 γ_Σ 最大。直观来看，信噪比高的支路应该有更大的权重，因此 a_i^2 应该正比于支路信噪比 r_i^2/N_0。通过对式（7-17）求偏导的方法或者通过柯西-许瓦兹不等式的方法，我们可以求解出能使 γ_Σ 最大化的最佳加权值正是 $a_i^2=r_i^2/N_0$。此时合并输出的信噪比为 $\gamma_\Sigma=\sum_{i=1}^{M}r_i^2/N_0=\sum_{i=1}^{M}\gamma_i$，即合并输出的信噪比是各支路信噪比的和。这样，合并输出的平均信噪比以及相应的阵列增益随分集支路数 M 的增加而线性增加，这一点与SC明显不同。与SC相同的一点是，即便各支路都是瑞利衰落，输出信噪比的分布也不是指数分布。

可利用矩母函数的乘积或特征函数的乘积来求得 γ_Σ 的分布。假设每个支路都是独立同分布的瑞利衰落，并有相同的平均支路信噪比 $\bar{\gamma}$，则 γ_Σ 的分布是自由度为 $2M$ 的卡方 (χ^2) 分布，其数学期望为 $\bar{\gamma}_\Sigma=M\bar{\gamma}$，方差为 $2M\bar{\gamma}$：

$$p_{\gamma_\Sigma}(\gamma)=\frac{\gamma^{M-1}e^{-\gamma/\bar{\gamma}}}{\bar{\gamma}^M(M-1)!}\quad \gamma\geqslant 0 \tag{7-18}$$

给定门限 γ_0 下相应的中断率为：

$$P_{\text{out}}=p(\gamma_\Sigma<\gamma_0)=\int_0^{\gamma_0}p_{\gamma_\Sigma}(\gamma)\,d\gamma=1-e^{-\gamma_0/\bar{\gamma}}\sum_{k=1}^{M}\frac{(\gamma_0/\bar{\gamma})^{k-1}}{(k-1)!} \tag{7-19}$$

图7-5示出了不同支路数下最大比合并的 P_{out}。

214

平均误码率可由式（7-2）求得。对于独立同分布瑞利衰落及BPSK调制，文献[4]的6.3节给出平均误码率为

$$\bar{P}_b = \int_0^\infty Q(\sqrt{2\gamma})p_{\gamma_\Sigma}(\gamma)\,\mathrm{d}\gamma = \left(\frac{1-\Gamma}{2}\right)^M \sum_{m=0}^{M-1}\binom{M-1+m}{m}\left(\frac{1+\Gamma}{2}\right)^m \tag{7-20}$$

式中 $\Gamma=\sqrt{\bar{\gamma}/(1+\bar{\gamma})}$，其曲线如图7-6所示。比较图7-5中MRC和图7-2中SC的中断率曲线或比较图7-6中MRC和图7-3中SC的误码率曲线可以看出，MRC的性能显著好于SC。7.4节中将用矩母函数的方法来分析MRC的平均误码率，该方法适用于任何调制方式、任意分集支路数或支路衰落分布。

215

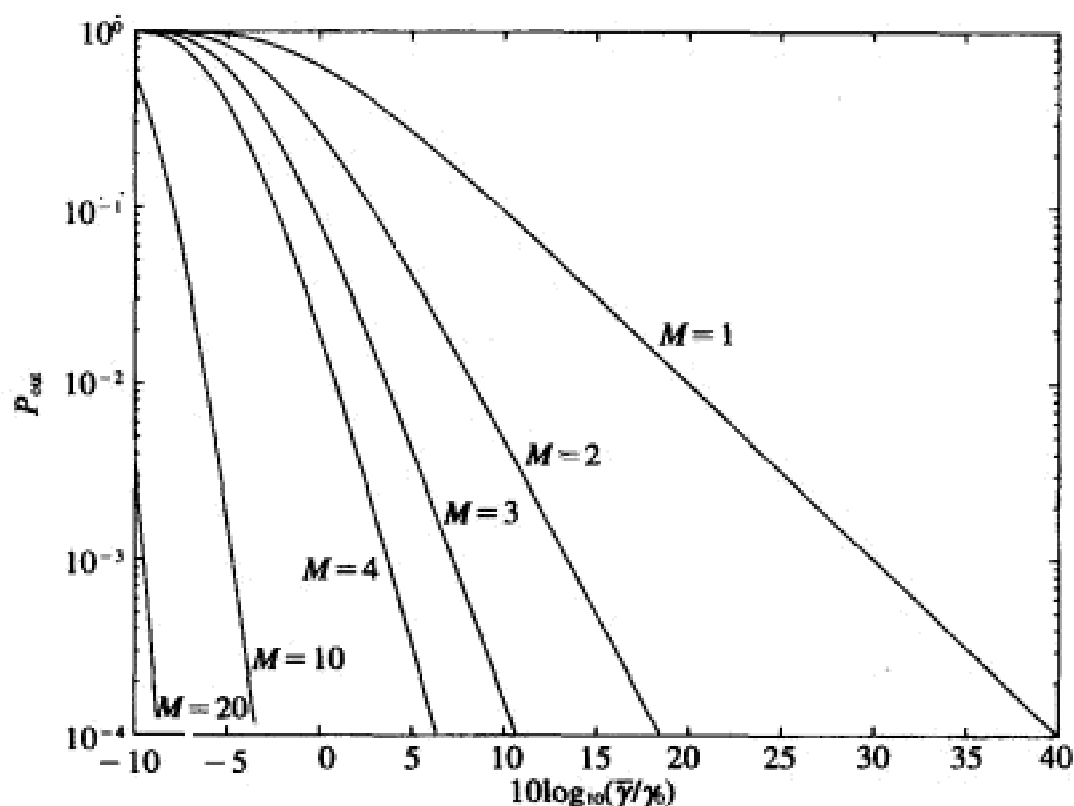


图7-5　独立同分布瑞利衰落下最大比合并的P_{out}

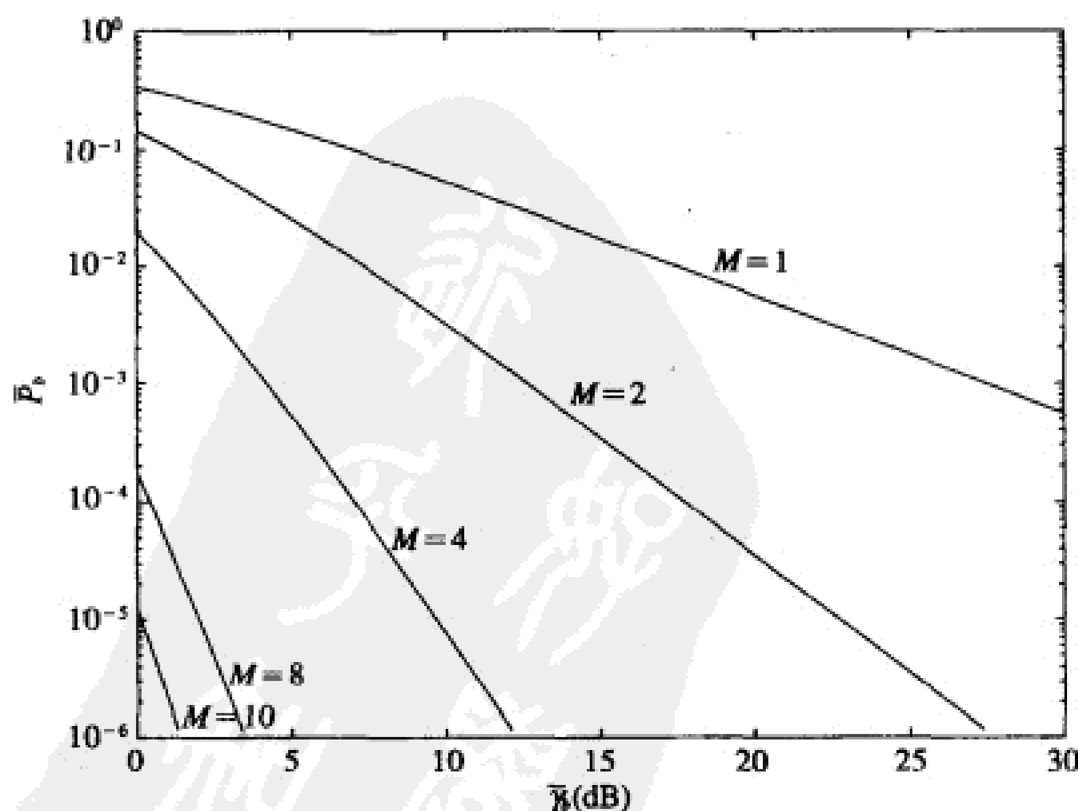


图7-6　独立同分布瑞利衰落下最大比合并的平均P_b

对Q函数应用切尔诺夫界$Q(x)\leqslant e^{-x^2/2}$，可得到一个简单的平均误码率上界。对于固定的信道增益，MRC的误码率近似为

$$P_s=\alpha_M Q\left(\sqrt{\beta_M\gamma_\Sigma}\right)\leqslant\alpha_M e^{-\beta_M\gamma_\Sigma/2}=\alpha_M e^{-\beta_M(\gamma_1+\cdots+\gamma_M)/2} \tag{7-21}$$

按卡方（χ^2）分布对γ_Σ积分得到

$$\bar{P}_s\leqslant\alpha_M\prod_{i=1}^{M}\frac{1}{1+\beta_M\bar{\gamma}_i/2} \tag{7-22}$$

假设γ_i同分布的，$\bar{\gamma}_i=\bar{\gamma}$，则在高信噪比条件下有

$$\bar{P}_s\approx\alpha_M\left(\frac{\beta_M\bar{\gamma}}{2}\right)^{-M} \tag{7-23}$$

即在高信噪比时MRC的分集阶数是M，等于天线数，所以MRC达到了满分集阶数。

7.2.5 等增益合并

最大比合并需要知道每个支路上的时变信噪比，这些值测量起来比较困难。简单一些的方法是*等增益合并*（equal-gain combining，EGC），它以相同的权重$\alpha_i=e^{-j\theta_i}$对各支路上的信号进行同相合并。假设每个支路上的噪声功率谱密度都是$N_0/2$，那么合并输出的信噪比为

$$\gamma_\Sigma=\frac{1}{N_0M}\left(\sum_{i=1}^{M}r_i\right)^2 \tag{7-24}$$

γ_Σ的概率密度函数和累积分布函数都不存在闭式表达。对于两支路分集，若支路上的衰落是独立同分布的瑞利分布，平均支路信噪比为$\bar{\gamma}$，则累积分布函数可以写出一个包含Q函数的表达式（见文献[4]的6.4节及文献[8]的5.6节）：

$$P_{\gamma_\Sigma}(\gamma)=1-e^{-2\gamma/\bar{\gamma}}-\sqrt{\pi\gamma/\bar{\gamma}}e^{-\gamma/\bar{\gamma}}\left(1-2Q\left(\sqrt{2\gamma/\bar{\gamma}}\right)\right) \tag{7-25}$$

相应得到中断率为

$$P_{\text{out}}(\gamma_0)=1-e^{-2\gamma_R}-\sqrt{\pi\gamma_R}e^{-\gamma_R}\left(1-2Q\left(\sqrt{2\gamma_R}\right)\right) \tag{7-26}$$

式中$\gamma_R=\gamma_0/\bar{\gamma}$。对式（7-25）中的$\gamma$微分可得概率密度函数为

$$p_{\gamma_\Sigma}(\gamma)=\frac{1}{\bar{\gamma}}e^{-2\gamma/\bar{\gamma}}-\sqrt{\pi}e^{-\gamma/\bar{\gamma}}\left(\frac{1}{\sqrt{4\gamma\bar{\gamma}}}-\frac{1}{\bar{\gamma}}\sqrt{\frac{\gamma}{\bar{\gamma}}}\right)\left(1-2Q\left(\sqrt{\frac{2\gamma}{\bar{\gamma}}}\right)\right) \tag{7-27}$$

216

代入式（7-2）可得到BPSK下的平均误比特率为

$$\bar{P}_b=\int_0^\infty Q\left(\sqrt{2\gamma}\right)p_{\gamma_\Sigma}(\gamma)\,d\gamma=0.5\left(1-\sqrt{1-\left(\frac{1}{1+\bar{\gamma}}\right)^2}\right) \tag{7-28}$$

文献[8]的5.7节指出EGC的性能非常接近MRC，功率损失不超过1dB。这个性能损失换来了等增益加权的低复杂度。更多关于SC、MRC和EGC的性能比较可以参看文献[1]的第9章。

例7.4：考虑两支路分集，每个支路是独立同分布的瑞利衰落，支路的平均信噪比为10dB。比较使用BPSK调制时，MRC和EGC的平均误码率。

解：由式（7-20），MRC的平均误码率为

$$\overline{P}_b=\left(\frac{1-\sqrt{10/11}}{2}\right)^2\left(2+\sqrt{10/11}\right)=1.60\times10^{-3}$$

由式（7-28），EGC的平均误码率为

$$\overline{P}_b=0.5\left(1-\sqrt{1-\left(\frac{1}{11}\right)^2}\right)=2.07\times10^{-3}$$

两者的性能几乎一样。

7.3 发送分集

发送分集中有多根发送天线，总发送功率是各天线上发送功率的和。发送分集适合于发送端在空间、供电能力及处理能力方面比接收端更富裕的系统，比如蜂窝系统。发送分集的设计与发送端是否知道复信道增益信息有关。当发送端已知信道复增益时，发送分集与接收分集非常类似。当发送端未知信道信息时，需要借助新近提出的Alamoui方案及其扩展，结合空时处理来获得发送分集增益。下面我们假设接收端已知信道增益，按发送端是否已知信道信息这两种情形分别讨论发送分集。

7.3.1 发送端已知信道信息

考虑有M根发送天线和一根接收天线的发送分集系统。假设发送端已知第i根天线上的路径增益$r_ie^{j\theta_i}$，称此为发送端已知信道边信息（channel side information，CSI），记为CSIT（CSI at the Transmitter）。记发送信号为$s(t)$，总发送的符号能量为E_s。信号乘以复增益$\alpha_i=a_ie^{-j\theta_i}(0\leqslant a_i\leqslant1)$后在第$i$根天线上发送。这个复数乘法同时实现同相和信道增益加权的功能。由于总发送能量是E_s，所以加权值必须满足$\sum_{i=1}^{M}a_i^2=1$。各天线发送的加权信号在空中相加，形成如下的接收信号

217

$$r(t)=\sum_{i=1}^{M}a_ir_is(t) \tag{7-29}$$

设接收端噪声的功率谱密度为$N_0/2$。我们希望通过设定合适的支路加权值使接收信噪比最大。按照接收MRC分集中类似的方法可得知，能使信噪比最大化的加权值a_i为

$$a_i=\frac{r_i}{\sqrt{\sum_{i=1}^{M}r_i^2}} \tag{7-30}$$

相应的信噪比为

$$\gamma_\Sigma=\frac{E_s}{N_0}\sum_{i=1}^{M}r_i^2=\sum_{i=1}^{M}\gamma_i \tag{7-31}$$

其中$\gamma_i=r_i^2E_s/N_0$是第i根发送天线和接收天线之间的支路信噪比。由此可见，当发送端

已知信道信息时，发送分集和接收MRC分集完全类似，其接收信噪比也是各支路信噪比的和。特别当各天线的路径增益相同为$r_i = r$时，$\gamma_\Sigma = Mr^2E_s/N_0$，采用$M$根发送天线时的信噪比是单天线时的$M$倍，即阵列增益为$M$。

给定各支路的路径增益时，类似7.2.4节中的做法，由切尔诺夫界可得

$$P_s = \alpha_M Q\left(\sqrt{\beta_M \gamma_\Sigma}\right) \leqslant \alpha_M e^{-\beta_M \gamma_\Sigma/2} = \alpha_M e^{-\beta_M(\gamma_1+\cdots+\gamma_M)/2} \tag{7-32}$$

按照卡方(χ^2)分布对γ_Σ进行积分，得到与式（7-22）相同的误码率上界

$$\bar{P}_s \leqslant \alpha_M \prod_{i=1}^{M} \frac{1}{1+\beta_M \bar{\gamma}_i/2} \tag{7-33}$$

假设γ_i分布相同，$\bar{\gamma}_i = \bar{\gamma}$，则在高信噪比条件下有

$$\bar{P}_s \approx \alpha_M \left(\frac{\beta_M \bar{\gamma}}{2}\right)^{-M} \tag{7-34}$$

因此高信噪比下MRC发送分集的分集阶数是M，所以发送和接收的MRC分集都达到了满分集阶数。发送端已知信道信息时，EGC和SC发送分集的相关分析同接收分集时一样。

发送分集的主要困难是要让发送端获得信道的相位和幅度信息。一种方法是，接收端利用导频测量信道，再将测量结果反馈给发送端。对于时分双工的蜂窝系统，基站可以测得移动台到基站的信道相位和幅度，再把这个测量结果作为基站向移动台发送时的信道信息，其原理是时分双工的系统中前反向信道有互易性。 218

7.3.2 发送端未知信道信息－Alamouti方案

考虑和7.3.1节相同的模型，但发送端并不知道信道增益$r_ie^{j\theta_i}$，也即无CSIT。这种情况下如何进行分集并不十分明显。考虑两根发送天线时的一种简单做法，假设每根天线的信道增益$h_i = r_ie^{j\theta_i}(i = 1, 2)$是均值为零、方差为1的复高斯随机变量，发送信号$s(t)$的符号能量为$E_s$。如果我们只是简单地平分功率后用两个天线发射，那么天线i上的发送信号将是$s_i(t) = \sqrt{0.5}s(t)$，接收信号为

$$r(t) = \sqrt{0.5}(h_1 + h_2)s(t) \tag{7-35}$$

两个复高斯随机变量的和$h_1 + h_2$本身也是复高斯的，和的均值为零、方差为2，因此$\sqrt{0.5}(h_1 + h_2)$仍然是均值为零、方差为1的复高斯随机变量，这和单根发送天线是一样的。可见这样的方法既不能使两个天线同相，也不能对两个天线加权分配能量，所以没有任何性能优点。

如果我们采用一些适当的天线技术，那么即使没有信道信息，也能获得发送分集增益。一种常用的特别简单的方法是由Alamouti发明的[9]，它结合了空间和时间分集。Alamouti方案是针对有两根发送天线的数字通信系统设计的。它占用两个符号周期并且假设在这段时间内信道增益不变。在第一个符号周期，两个不同的符号s_1和s_2分别用天线1和天线2同时发送，每个符号的能量为$E_s/2$，在下一个符号周期，天线1发送$-s_2^*$，天

线2发送 s_1^*，每个的符号能量同样也为$E_s/2$。

假设第i根发送天线和接收天线之间的复信道增益为$h_i = r_i e^{j\theta_i}(i=1, 2)$。第一个符号周期的接收信号为$y_1 = h_1 s_1 + h_2 s_2 + n_1$，第二个符号周期的接收信号为$y_2 = -h_1 s_2^* + h_2 s_1^* + n_2$，$n_i(i=1, 2)$是第$i$个发送周期内接收端的加性高斯白噪声的样值，假设噪声样值的均值为零、功率为N_0。

这两个先后接收到的符号形成矢量$\boldsymbol{y} = \left[y_1\, y_2^*\right]^T$，可写为

$$\boldsymbol{y} = \begin{bmatrix} h_1 & h_2 \\ h_2^* & -h_1^* \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} + \begin{bmatrix} n_1 \\ n_2^* \end{bmatrix} = \boldsymbol{H}_A \boldsymbol{s} + \boldsymbol{n}$$

式中$\boldsymbol{s} = [s_1, s_2]^T$，$\boldsymbol{n} = [n_1, n_2]^T$，$\boldsymbol{H}_A = \begin{bmatrix} h_1 & h_2 \\ h_2^* & -h_1^* \end{bmatrix}$。$\boldsymbol{H}_A$的结构满足

219

$$\boldsymbol{H}_A^H \boldsymbol{H}_A = (|h_1^2| + |h_2^2|)\boldsymbol{I}_2 \tag{7-36}$$

这是一个对角阵，因此若定义一个新向量$\boldsymbol{z} = \boldsymbol{H}_A^H \boldsymbol{y}$，则有

$$\boldsymbol{z} = [z_1\, z_2]^T = (|h_1^2| + |h_2^2|)\boldsymbol{I}_2 \boldsymbol{s} + \tilde{\boldsymbol{n}} \tag{7-37}$$

式中$\tilde{\boldsymbol{n}} = \boldsymbol{H}_A^H \boldsymbol{n}$是复高斯噪声矢量，其均值为零，协方差矩阵为$E[\tilde{\boldsymbol{n}}\tilde{\boldsymbol{n}}^*] = (|h_1^2| + |h_2^2|)N_0 \boldsymbol{I}_2$。$\boldsymbol{z}$表达式中的对角阵特性使我们可以分离出两个符号的传输，z的每个元素对应于一个符号的发送：

$$z_i = (|h_1^2| + |h_2^2|)s_i + \tilde{n}_i \qquad i = 1, 2 \tag{7-38}$$

对应z_i的接收信噪比为

$$\gamma_i = \frac{(|h_1^2| + |h_2^2|)E_s}{2N_0} \tag{7-39}$$

其中的因子2源于发送信号s_i的能量是总符号能量E_s的一半。接收信噪比等于各支路信噪比的和除以2。由式（7-38）可知，虽然发送端并不知道信道信息，Alamouti方案的分集阶数仍可达到2，是两天线发送分集可能达到的最大值。但从式（7-39）看，阵列增益只能达到1，而MRC的阵列增益和分集增益都是2。

Alamouti方案推广到$M>2$的情形，推广后的设计属于正交空时分组码，详见文献[10]的6.3.3节及文献[11]的7.4节。

7.4 利用矩母函数分析分集

本节利用6.3.3节中介绍的矩母函数大大简化分集系统中平均误码率的分析。由于求解合并输出信噪比γ_Σ的分布$p_{\gamma_\Sigma}(\gamma)$有相当的困难，所以引入了矩母函数。虽然式（7-2）和式（7-3）给出的平均误码率和中断率计算公式很简单，但它们需要对γ_Σ的概率密度函数$p_{\gamma_\Sigma}(\gamma)$进行积分。无论采用何种合并方式，对于任意的支路数及各支路分布不同的情形，$p_{\gamma_\Sigma}(\gamma)$往往没有闭式解。特别是，$p_{\gamma_\Sigma}(\gamma)$往往是无穷积分的形式，这使式（7-2）和式（7-3）变成了难以数值求解的双重积分。即便$p_{\gamma_\Sigma}(\gamma)$有闭式解，式（7-2）和式（7-3）的积分也不一定有闭式解，数值积分也有可能比较困难。对于不同调制技术、不同衰落和不同合并技术下的平均误码率积分问题，过去几十年中有大量的工作是关于如

何近似这个积分或者如何进行数值求解（见文献[12]及其中的参考文献）。用矩母函数替代概率密度函数来表示平均误码率，往往可以避免复杂的积分。特别当分集衰落路径独立但不一定同分布时，这种方法一般能给出平均误码率的闭式解，或是能给出一个容易进行数值求解的有限区间上的单重积分。

在分集系统的性能分析中，矩母函数方法最简单的应用就是用于MRC合并的相干调制。因此我们首先讨论这个问题。然后再讨论矩母函数在EGC和SC平均误码率分析中的应用。

220

7.4.1 MRC分集的分析

从7.2.4节我们知道，最大比合并输出的信噪比 γ_Σ 是各支路信噪比γ_i之和：

$$\gamma_\Sigma = \sum_{i=1}^{M} \gamma_i \tag{7-40}$$

这一点使得矩母函数的方法非常简便。如同6.3.3节中分析无分集系统平均误码率时的情形一样，我们仍然假设所研究调制方式在AWGN信道下的误码率可以表示成γ_i的指数函数（6-67）或有限积分式（6-68）。

先考虑P_s为式（6-67）的形式，此时MRC下的平均误码率为

$$\bar{P}_s = \int_0^\infty c_1 \exp[-c_2\gamma] p_{\gamma_\Sigma}(\gamma)\,\mathrm{d}\gamma \tag{7-41}$$

假设各支路的信噪比独立，则联合概率密度函数是各自概率密度函数之积$p_{\gamma_1,\cdots,\gamma_M}(\gamma_1,\cdots,\gamma_M)=p_{\gamma_1}(\gamma_1)\cdots p_{\gamma_M}(\gamma_M)$。将这个因式分解及$\gamma=\gamma_1+\cdots+\gamma_M$代入式（7-41）得到

$$\bar{P}_s = c_1 \underbrace{\int_0^\infty \int_0^\infty \cdots \int_0^\infty}_{M\text{-层}} \exp[-c_2(\gamma_1+\cdots+\gamma_M)] p_{\gamma_1}(\gamma_1)\ldots p_{\gamma_M}(\gamma_M)\,\mathrm{d}\gamma_1\ldots\mathrm{d}\gamma_M \tag{7-42}$$

再将乘积形式 $\exp[-c_2(\gamma_1+\cdots+\gamma_M)]=\prod_{i=1}^{M}\exp[-c_2\gamma_i]$ 和 $p_{\gamma_1}(\gamma_1)\cdots p_{\gamma_M}(\gamma_M)=\prod_{i=1}^{M}p_{\gamma_i}(\gamma_i)$ 代入到式（7-42）中得到：

$$\bar{P}_s = c_1 \underbrace{\int_0^\infty \int_0^\infty \cdots \int_0^\infty}_{M\text{-层}} \prod_{i=1}^{M} \exp[-c_2\gamma_i] p_{\gamma_i}(\gamma_i)\,\mathrm{d}\gamma_i \tag{7-43}$$

交换式（7-43）中积分和乘积的次序，便得到我们期望的最终形式

$$\bar{P}_s = c_1 \prod_{i=1}^{M} \int_0^\infty \exp[-c_2\gamma_i] p_{\gamma_i}(\gamma_i)\,\mathrm{d}\gamma_i = c_1 \prod_{i=1}^{M} \mathcal{M}_{\gamma_i}(-c_2) \tag{7-44}$$

式中$M_{\gamma_i}(s)$是i支路衰落的矩母函数。这样，平均误符号率是各支路信噪比的矩母函数之积。对于瑞利、莱斯及Nakagami衰落，矩母函数分别由式（6-63）、式（6-64）和式（6-65）给出。

当P_s具有式（6-68）的形式时，类似有：

$$\bar{P}_s = \int_0^\infty \int_A^B c_1 \exp[-c_2(x)\gamma]\,\mathrm{d}x\, p_{\gamma_\Sigma}(\gamma)\,\mathrm{d}\gamma$$
$$= \underbrace{\int_0^\infty \int_0^\infty \cdots \int_0^\infty}_{M\text{-层}} \int_A^B c_1 \prod_{i=1}^{M} \exp[-c_2(x)\gamma_i] p_{\gamma_i}(\gamma_i)\,\mathrm{d}\gamma_i \tag{7-45}$$

221

交换积分和乘积的次序得到

$$\bar{P}_s = c_1 \int_A^B \prod_{i=1}^{M} \int_0^\infty \exp[-c_2(x)\gamma_i] p_{\gamma_i}(\gamma_i)\,\mathrm{d}\gamma_i = c_1 \int_A^B \prod_{i=1}^{M} \mathcal{M}_{\gamma_i}(-c_2(x))\,\mathrm{d}x \tag{7-46}$$

平均误码率是各支路信噪比的矩母函数乘积的有限一重积分。式（7-44）和式（7-46）非常简单，只要求各支路的信噪比相互独立，适用于任意的分集支路数和任意的支路衰落分布。

现在我们来看具体调制和具体衰落分布时的平均误码率。首先考虑DPSK调制，它在AWGN信道中的误码率是$P_b(\gamma_b)=0.5\mathrm{e}^{-\gamma_b}$，具有式（6-67）的形式，$c_1=1/2$、$c_2=1$。于是，DPSK在M重MRC分集下的平均误码率可根据式（7-44）得到：

$$\bar{P}_b = \frac{1}{2}\prod_{i=1}^{M} \mathcal{M}_{\gamma_i}(-1) \tag{7-47}$$

$M=1$时就是式（6-60）所给出的无分集时的平均误码率。

例7.5：计算三支路MRC分集下DPSK的平均误码率。假设各支路是独立的瑞利衰落，平均信噪比是$\bar{\gamma}_1=15\text{dB}$、$\bar{\gamma}_2=\bar{\gamma}_3=5\text{dB}$。并与平均信噪比为15dB的无分集情形做比较。

解：由式（6-63），$\mathcal{M}_{\gamma_i}(s)=(1-s\bar{\gamma}_i)^{-1}$。代入$s=-1$，由式（7-47）得

$$\bar{P}_b = \frac{1}{2}\times\frac{1}{1+10^{1.5}}\times\left(\frac{1}{1+10^{0.5}}\right)^2 = 8.85\times10^{-4}$$

无分集时

$$\bar{P}_b = \frac{1}{2}\times\frac{1}{1+10^{1.5}} = 1.53\times10^{-2}$$

222

可见，即使增加的支路信噪比很低，增加分集支路也能大大降低平均误码率。

例7.6：计算三分支MRC分集下DPSK的平均误码率。假设第一支路为Nakagami分布，$m=2$，$\bar{\gamma}_1=15\text{dB}$；第二支路为莱斯分布，$K=3$，$\bar{\gamma}_2=5\text{dB}$；第三支路为Nakagami分布，$m=4$，$\bar{\gamma}_3=5\text{dB}$。并与上例的结果进行比较。

解：由式（6-64）和式（6-65）可得，Nakagami分布的矩母函数是$\mathcal{M}_{\gamma_i}(s)=(1-s\bar{\gamma}_i/m)^{-m}$，莱斯分布的矩母函数是

$$\mathcal{M}_{\gamma_s}(s) = \frac{1+K}{1+K-s\bar{\gamma}_s}\exp\left[\frac{Ks\bar{\gamma}_s}{1+K-s\bar{\gamma}_s}\right]$$

代入$s=-1$，由式（7-47）得

$$\bar{P}_b = \frac{1}{2}\left(\frac{1}{1+10^{1.5}/2}\right)^2 \frac{4}{4+10^{0.5}}\exp\left[\frac{-3\times10^{0.5}}{4+10^{0.5}}\right]\left(\frac{1}{1+10^{0.5}/4}\right)^4 = 6.9\times10^{-5}$$

本例和上例相比，各支路都是独立衰落，每个支路的平均信噪比也相同，但本例的平均误码率降低了一个数量级还要多。说明Nakagami分布和莱斯分布优于瑞利分布，特别是对多支路的MRC合并。从本例可以看到使用矩母函数的优越性。当各支路信噪比的分布不同时，用矩母函数的方法计算平均误码率，结果是一些不同函数的乘积组成的简单闭式解。相反，求不同分布的随机变量的和的概率密度函数需要用到这些概率密度函数的卷积，这个卷积一般没有闭式结果。

对BPSK调制，从式（6-44）可见，P_b具有与式（6-68）式相同的形式[0]，其中$x=\phi$、$c_1=1/\pi$、$A=0$、$B=\pi/2$、$c_2(\phi)=1/\sin^2\phi$。于是，BPSK在M重分集下的平均误比特率为

$$\bar{P}_b=\frac{1}{\pi}\int_0^{\pi/2}\prod_{i=1}^{M}\mathcal{M}_{\gamma_i}\left(-\frac{1}{\sin^2\phi}\right)\mathrm{d}\phi \tag{7-48}$$

类似地，若$P_s=\alpha Q\left(\sqrt{2g\gamma_s}\right)$，那么$P_s$就具有式（6-68）形式，其中$x=\phi$、$c_1=\alpha/\pi$[1]、$A=0$、$B=\pi/2$、$c_2(\phi)=g/\sin^2\phi$。$M$重分集下的平均误码率为

$$\bar{P}_s=\frac{\alpha}{\pi}\int_0^{\pi/2}\prod_{i=1}^{M}\mathcal{M}_{\gamma_i}\left(-\frac{g}{\sin^2\phi}\right)\mathrm{d}\phi \tag{7-49}$$

如果各支路的信噪比为独立同分布，则上式简化为

$$\bar{P}_s=\frac{\alpha}{\pi}\int_0^{\pi/2}\left(\mathcal{M}_{\gamma}\left(-\frac{g}{\sin^2\phi}\right)\right)^M\mathrm{d}\phi \tag{7-50}$$

式中$\mathcal{M}_\gamma(s)$是支路信噪比的矩母函数，各支路的矩母函数是相同的。

从式（6-45）可以看出MPSK的误码率也具有式（6-68）的形式，因此平均误码率为

$$\bar{P}_s=\frac{1}{\pi}\int_0^{(M-1)\pi/M}\prod_{i=1}^{M}\mathcal{M}_{\gamma_i}\left(-\frac{g}{\sin^2\phi}\right)\mathrm{d}\phi \tag{7-51}$$

其中$g=\sin^2\left(\frac{\pi}{M}\right)$。独立同分布时简化为：

$$\bar{P}_s=\frac{1}{\pi}\int_0^{(M-1)\pi/M}\left(\mathcal{M}_{\gamma}\left(-\frac{g}{\sin^2\phi}\right)\right)^M\mathrm{d}\phi \tag{7-52}$$

223

例7.7：求二支路MRC合并下8PSK调制的平均误码率表达式。假设每个支路都是平均信噪比为20dB的瑞利衰落。

解：瑞利衰落的矩母函数为$\mathcal{M}_{\gamma_i}(s)=(1-s\bar{\gamma}_i)^{-1}$。在式（7-52）中代入此矩母函数及$s=-(\sin^2\pi/8)/\sin^2\phi$及$\bar{\gamma}=100$得到

$$\bar{P}_s=\frac{1}{\pi}\int_0^{7\pi/8}\left(\frac{1}{1+(100\sin^2\pi/8)/\sin^2\phi}\right)^2\mathrm{d}\phi$$

这个表达式没有闭式解，只能进行数值求解，结果为$\bar{P}_s=1.56\times10^{-3}$。

1. 作者的勘误表指出，原书$c_1=1/\pi$错误，应为$c_1=\alpha/\pi$，此处已改。——译者注

也可以用类似的方法将MQAM在衰落信道下的误码率式（6-81）推广到MRC分集的情形。具体是，首先将AWGN信道下的P_s用Q和Q^2的等效形式表示成式（6-80），再按γ_Σ的分布对其积分。由于$\gamma_\Sigma=\sum_i\gamma_i$，并且信噪比是独立的，所以对式（6-80）积分时，指数函数以及γ_Σ的概率密度函数都能写成乘积形式。然后类似于MPSK中的推导，交换积分和相乘的次序，就可以得到MQAM在MRC合并下的平均误码率为

$$\begin{aligned}\bar{P}_s=&\frac{4}{\pi}\left(1-\frac{1}{\sqrt{M}}\right)\int_0^{\pi/2}\prod_{i=1}^{M}\mathcal{M}_{\gamma_i}\left(-\frac{g}{\sin^2\phi}\right)\mathrm{d}\phi\\&-\frac{4}{\pi}\left(1-\frac{1}{\sqrt{M}}\right)^2\int_0^{\pi/4}\prod_{i=1}^{M}\mathcal{M}_{\gamma_i}\left(-\frac{g}{\sin^2\phi}\right)\mathrm{d}\phi\end{aligned}\tag{7-53}$$

对于更为广泛的调制类型，利用矩母函数求解M重MRC分集下的平均误码率问题可参见文献[1]的9.2节。

7.4.2　EGC和SC分集的分析

矩母函数的方法在EGC和SC的性能分析中不如在MRC中好用。原因在于，MRC中$\gamma_\Sigma=\sum_i\gamma_i$并且$\exp[-c_2\gamma_\Sigma]=\prod_i\exp[-c_2\gamma_i]$，这一点直接简化了公式，它使平均误码率成为各个支路信噪比矩母函数的积。而EGC和SC中没有这种关系。不过在某些情况下，用γ_Σ的矩母函数进行分析仍然要比直接用分布进行分析更为简单。比如文献[1]在9.3.3节中用各支路信噪比的特征函数得到了EGC分集时MPSK误码率的精确表达式。特征函数是矩母函数在$s=\mathrm{j}2\pi f$时的结果，也即γ_Σ的概率密度函数的傅里叶变换。文献[1]中的式（9-78）给出的平均误码率是一些闭式表达式在有限区间上的积分，很容易进行数值计算。

7.4.3　非相干及差分相干调制时分集系统的性能分析

也可以用类似的方法用矩母函数来分析非相干调制和差分相干调制在分集系统中的误码率，文献[1]和[13]中给出了这种方法。非相干调制和差分相干调制与相干调制时的区别是，AWGN信道下非相干调制和差分相干调制的误码率是用Marcum Q函数的形式给出的，而不是高斯Q函数，这导致用矩母函数分析时存在差别，除此之外的推导分析并无实质差别。所得到的误码率表达式是一个有限区间的一重积分，容易进行数值计算，详见文献[1]和[13]。

224

习题

7.1　考虑瑞利衰落信道的SC分集，求支路数为$M=1$（无分集）、$M=2$及$M=3$时QPSK对应于$P_s=10^{-3}$的中断率。假设这三个支路的平均信噪比分别为$\bar{\gamma}_1=10\text{dB}$、$\bar{\gamma}_2=15\text{dB}$及$\bar{\gamma}_3=20\text{dB}$。

7.2　式（7-9）给出了瑞利衰落信道下，M支路选择合并的输出信噪比的概率密度函数$p_{\gamma_\Sigma}(\gamma)$。假设每个支路的平均信噪比为10dB，请画出$M=1$、2、4、8、10时的$p_{\gamma_\Sigma}(\gamma)$曲线。要求横坐标和纵坐标都采用线性刻度，$\gamma$的范围为$0\leqslant\gamma\leqslant60$。讨论$M$增加时分布的变化，

以及它为什么能使误码率降低。

7.3　请推导出DPSK在独立同分布瑞利衰落及SC下的平均误码率公式，即式（7-11）。

7.4　对于各支路并非独立同分布的情形，推导出两支路SSC输出信噪比的累积概率分布函数的一般表达式，并证明它对于独立同分布的情形能退化为式（7-12）。并就每个支路为不同平均信噪比的瑞利衰落，给出相应的具体表达式。

7.5　请推导DPSK在独立同分布瑞利衰落及SSC下的平均误码率公式，即式（7-16）。

7.6　画出DPSK在MRC下的平均误码率曲线。假设支路数$M=2$、3、4，各支路为独立同分布的瑞利衰落，平均支路信噪比的范围为0~20dB。

7.7　证明MRC接收分集中，能使γ_Σ最大化的加权系数a_i为$a_i^2=r_i^2/N_0$，其中$N_0/2$是各支路的噪声功率谱密度。同时证明，在该加权系数下，$\gamma_\Sigma=\sum_i\gamma$。

7.8　本题说明，由于阵列增益的原因，即使没有衰落，分集合并也能带来性能增益。考虑N支路的分集合并系统，每个支路是信噪比为$\gamma_i=10$dB的AWGN信道。假设采用$M=4$的MQAM调制，其误码率近似为$P_b=0.2e^{-1.5\gamma/(M-1)}$，其中$\gamma$是接收信噪比。

（a）求$N=1$时的P_b。

（b）MRC下，求使$P_b<10^{-6}$的N。

7.9　请推导BPSK在独立同分布瑞利衰落及MRC下的平均误码率公式，即式（7-20）。

7.10　请推导BPSK在独立同分布瑞利衰落及EGC下的平均误码率公式，即式（7-28）。　225

7.11　假设两支路分集系统中每个支路上的衰落是独立同分布的瑞利衰落。请比较BPSK调制在无分集、两支路SC、$\gamma_T=\gamma_0$的两支路SSC、两支路EGC以及两支路MRC下的平均误码率。考虑支路信噪比都是10dB和都是20dB的情形，请说明支路信噪比增加时，这些分集方式之间的相对性能有何差别？

7.12　假设两支路分集系统中每个支路上的衰落是独立同分布的瑞利衰落。请画出BPSK在MRC和EGC下的平均误码率。考虑支路平均信噪比的范围为0~20dB。请问EGC与MRC相比，性能损失最大是多少dB？

7.13　假设两支路分集系统中每个支路上的衰落是独立同分布的瑞利衰落，支路的平均信噪比是$\bar{\gamma}=10$dB。假设采用BPSK调制，可接受误码率为$P_b=10^{-3}$。请比较MRC和EGC的中断率。

7.14　假设两支路分集系统中每个支路上的衰落是独立同分布的瑞利衰落，支路的平均信噪比是$\bar{\gamma}=10$dB。比较BPSK调制在MRC和EGC下的平均误码率。

7.15　请计算Alamouti方案在两支路发送分集下的平均误比特率，假设支路的平均信噪比为10dB。

7.16　已知某二支路MRC分集系统中两个支路上的衰落是独立同分布的，每支路的平均信噪比为10dB，每支路上衰落的分布均为$p(\gamma)$且$p(\gamma)$满足$\int_0^\infty p(\gamma)e^{-x\gamma}d\gamma=0.01\bar{\gamma}/\sqrt{x}$。求BPSK的平均$P_b$。

7.17　已知某三支路MRC分集系统中各个支路上的衰落是独立同分布的，每支路的平均信噪比为15dB。请计算衰落类型分别为瑞利分布和$m=2$的Nakagami分布时，BPSK的平均误比特率。提示：可利用Q函数的等效形式来简化计算，至少Nakagami衰落时需要如此。

7.18　请画出BPSK在二支路MRC下的平均误码率作为支路信噪比函数的曲线。假设第

一条支路是瑞利衰落，第二条支路是$m=2$的Nakagami-m衰落，两个支路的平均信噪比相同。要求图中支路信噪比的范围为5dB~20dB。

7.19　假设M支路MRC系统中，每个分支上都是独立同分布的瑞利衰落，请就$M=1$、2、4、8的情形，画出8PSK的平均误码率作为支路信噪比函数的曲线。要求支路信噪比的范围是5dB~20dB。

7.20　MQAM在AWGN信道下的误码率可用Q及Q^2函数的等效表达式写成式（6-80）。请据此推导出MQAM在MRC下的平均误码率公式，即式（7-53）。

7.21　考虑三支路的MRC分集系统，假设第一个支路是瑞利衰落，第二、三个支路是226$K=2$的莱斯衰落，各支路的平均信噪比都是10dB。请比较16PSK和16QAM的平均误码率。

7.22　假设M支路MRC系统中，每个分支上都是独立同分布的瑞利衰落，请就$M=1$、2、4、8的情形，画出16QAM调制下的平均误码率与支路信噪比关系的曲线。要求支路信噪比的范围是5dB~20dB。

参考文献

[1] M. K. Simon and M.-S. Alouini, *Digital Communication over Fading Channels: A Unified Approach to Performance Analysis*, Wiley, New York, 2000.
[2] W. C. Y. Lee, *Mobile Communications Engineering*, McGraw-Hill, New York, 1982.
[3] J. Winters, "Signal acquisition and tracking with adaptive arrays in the digital mobile radio system IS-54 with flat fading," *IEEE Trans. Veh. Tech.*, pp. 1740–51, November 1993.
[4] G. L. Stuber, *Principles of Mobile Communications*, 2nd ed., Kluwer, Dordrecht, 2001.
[5] M. Blanco and K. Zdunek, "Performance and optimization of switched diversity systems for the detection of signals with Rayleigh fading," *IEEE Trans. Commun.*, pp. 1887–95, December 1979.
[6] A. Abu-Dayya and N. Beaulieu, "Switched diversity on microcellular Ricean channels," *IEEE Trans. Veh. Tech.*, pp. 970–6, November 1994.
[7] A. Abu-Dayya and N. Beaulieu, "Analysis of switched diversity systems on generalized-fading channels," *IEEE Trans. Commun.*, pp. 2959–66, November 1994.
[8] M. Yacoub, *Principles of Mobile Radio Engineering*, CRC Press, Boca Raton, FL, 1993.
[9] S. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Sel. Areas Commun.*, pp. 1451–8, October 1998.
[10] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*, Cambridge University Press, 2003.
[11] E. G. Larsson and P. Stoica, *Space-Time Block Coding for Wireless Communications*, Cambridge University Press, 2003.
[12] M. K. Simon and M.-S. Alouini, "A unified approach to the performance analysis of digital communications over generalized fading channels," *Proc. IEEE*, pp. 1860–77, September 1998.
[13] M. K. Simon and M.-S. Alouini, "A unified approach for the probability of error for noncoherent and differentially coherent modulations over generalized fading channels," *IEEE Trans. Commun.*, pp. 1625–38, December 1998.

227

第8章　无线信道中的编码

通过编码，无线信道中的比特差错可以通过接收端的译码器检测或者纠正。可以认为，编码是将信号星座点的维数进行了扩展。维数的扩展增加了星座点之间的距离，使我们能够更好地检测出错误或者纠正错误。

本章介绍针对AWGN信道及衰落信道而设计的编码。针对AWGN信道所设计的码一般不能很好的工作在衰落信道中，因为它们不能纠正深衰落造成的长串突发错误。用于衰落信道的编码一般是在AWGN信道编码的基础上结合使用交织器。不过为了获得分集效果，衰落信道中编码的设计准则有所改变。此外，抗衰落编码技术还包括不等错误保护编码及信源信道联合编码。

本章首先概述衰落信道和AWGN信道中的编码设计，介绍相关的一些基本设计参数，如最小距离、编码增益、带宽扩展、分集阶数等。8.2节和8.3节对AWGN信道中的分组码（block code）和卷积码（convolutional code）进行简要叙述。这些码虽然没有直接用在衰落信道上，但衰落信道中的纠错编码以及无线通信中使用的其他码（如CDMA系统中使用的扩频码）需要这些基础背景。本章也将介绍AWGN信道中的级联码（concatenated code）及由其演进而来的Turbo码和LDPC（low-density parity-check）码。这些高效的编码能以合理的复杂度接近香农极限。20世纪70年代末期发明了编码与调制联合设计的编码调制（coded modulation）技术，我们将介绍网格（trellis）编码调制以及更一般化的点格（lattice）编码调制的基本原理，以及它们在AWGN信道下的性能。

8.8节讨论衰落信道下的编码设计，这些编码是AWGN信道下的分组码或卷积码与交织的结合，同时，编码的设计准则变成了分集最大化。比特或符号交织的编码调制也可以获得分集增益，其中比特交织的分集增益通常更高。交织编码和其他形式的分集技术一样有分集增益，其分集阶数由编码的内在设计决定。为对抗信道衰落，不等错误保护也是分集的一个很好替代。不同于分集，不等错误保护将比特分为不同的优先级，对高优先级的比特采用更强的编码以对抗深衰落。由于比特优先级也是信源编码的一部分，所以不等差错保护实际上就是后面将要讨论的信源信道联合编码的一个特例。

228

编码是一门内容宽广深厚的学科，有许多优秀的书籍专门介绍这方面的知识。本章假定读者没有学过编码，因此先对AWGN信道中的编码进行了一定深度的介绍，然后再讨论无线系统中的编码设计。如果读者只是想粗略了解无线信道中的编码，只需专注于8.1节和8.8节。

8.1　码设计概述

无线通信系统采用纠错编码的主要目的是为了降低比特或者数据帧的错误概率。编码系统中的比特错误概率P_b指译码后出现比特错误的概率，帧错误率（分组错误率）P_{bl}指由许多比特组成的数据块经过译码后存在一个或多个错误比特的概率。分组数据系统

中的数据是按块的方式发送的，此时帧错误率指标更为合适。特定编码对差错数量的降低程度由AWGN信道中的编码增益和衰落信道中的分集增益反映。

AWGN信道中的编码增益定义为给定P_b或者P_{bl}时，采用编码所需要的比特能量或者信噪比相对于无编码时的减小量。如图8-1所示，$P_b=10^{-4}$时的增益C_{g1}要比$P_b=10^{-6}$时的增益C_{g2}小，而在$P_b=10^{-2}$时几乎没有编码增益。实际上，为高信噪比信道设计的码在低信噪比时有可能产生负的编码增益，这是因为编码后信息比特的能量扩散在多个编码比特上，每个编码比特的能量有所降低，如果信噪比很低的话，编码冗余带来的性能增益可能弥补不了编码比特能量的降低。因此，信道中信噪比的波动会显著降低编码的性能。

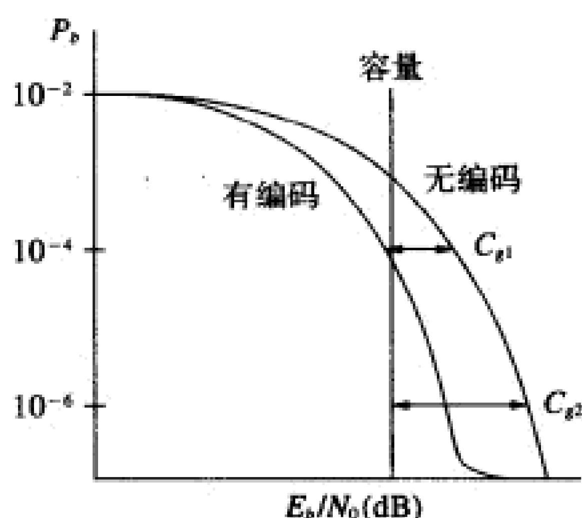


图8-1　AWGN信道下的编码增益

AWGN信道中的编码增益一般与该码的最小欧氏距离有关，最小欧氏距离是信号空间中的码字或错误事件之间的最小距离。为了获得好的性能，AWGN信道中的编码设计应当尽量使欧氏距离最大化。

229

无论有无编码，中低信噪比时错误率一般都随信噪比的增加而呈现出瀑布型下降。但某些编码（如Turbo码）在信噪比增加到一定程度后，会出现背景误码（error floor）现象。如图8-1所示，编码系统的错误率下降趋势在某个门限信噪比时突然发生变化，这个门限信噪比的数值与码的设计有关。信噪比高于门限时，码的性能主要取决于最小距离的错误事件，这个因素造成错误率曲线的下降斜率变小。

经常也通过和信道容量进行对比来衡量码的性能。图8-1中的直线标出了系统所传速率等于信道容量$B\log_2(1+\text{SNR})$时对应的信噪比或E_b/N_0值。对于可接近容量的编码，此时的误码率将趋于零。容量曲线反映实际编码所能获得的最好性能。

编码的纠错能力一般都是有代价的，这个代价可能是增加了编码的复杂度，如分组码、卷积码、Turbo码、LDPC码等，或者是降低了数据率、增加了信号带宽。假设一种编码把k个信息比特编为n个编码比特，那么它实际是把一个k维子空间扩展到了一个更大的n维空间中，从而使码字间有更大的距离。如果信道的数据速率固定为R_b，那么采用这种编码时的信息速率是$(k/n)R_b$，即编码使数据率降低到为原来的k/n。也可以把发送编码比特的周期降低到原来的k/n从而保持信息速率不变并能获得编码增益，但这将使发送信号的带宽展宽为原来的n/k倍。8.7节所讨论的编码调制则采用编码和调制联合设计的方法，不需要扩展带宽就能获得编码增益。

衰落信道存在不能纠正的突发错误，所以为AWGN信道设计的编码在衰落信道中往往不能良好工作。为了获得好的性能，可将AWGN信道中的编码和交织结合起来使用，同时在码设计的时候考虑到优化分集特性。交织器把突发错误在时间域上打散，形成了一种时间分集，这种时间分集由编码的内在特性来实现。其性能和MRC分集类似，分集阶数等于码的最小汉明距离，即两个码字（或两个错误事件）中不相同的符号个数。

衰落信道中交织编码的设计原则是让汉明距离最大化。 230

8.2 线性分组码

线性分组码在概念上来说是单比特校验码的扩展。单比特检验码是一种常见的检错码。它在一个n比特的数据分组中加入一个比特，以指示该分组中1的数目是奇数还是偶数。如果发生了单个比特错误，不论错在校验位还是信息位，结果都将使信息比特中1的个数和校验比特的指示不一致，这样接收端就知道发生了错误。线性分组码拓展了这种思想，为了能够检测多个错误或者能够纠正一个或多个错误，它使用了更多的校验比特。不过我们将会看到，线性分组码以及卷积码的检错纠错能力是以带宽扩展或者数据率降低为代价的。以下将只考虑二进制编码，其输入信息及编码结果均由二进制比特0或者1组成。

8.2.1 二进制线性分组码

(n, k) 二进制分组码是由k个信息比特生成n个编码比特，编码比特也称为码字符号。对应所有n个比特的各种组合，n个编码比特有2^n个可能的取值。从中选出2^k个作为码字，将每个k比特的信息组同这2^k个码字一一对应，就形成了二进制分组码。该码的编码率是$R_c=k/n$，表示每码字符号所携带的信息比特数。若码字符号以每秒R_s个符号的速率传输，那么(n, k)分组码的信息速率就是$R_b=R_cR_s=(k/n)R_s$bit/s。可见分组码使数据率降低为无编码时的R_c倍。

当k个信息比特到n个码字符号的映射是线性映射时，这样的分组码称为线性码。为了更好地描述这种映射及相关的编译码，我们先来讨论二进制n维向量构成的向量空间及其子空间的特性。由所有二进制n维向量构成的集合B_n是二元域上的向量空间。二元域只包括0、1这两个元素。所有的域都有加法和乘法这两种运算。二元域的加法就是模2加，二元域的乘法就是普通乘法。如果B_n的子集S满足以下条件，则称它为B_n的子空间：

(1) S中包含全零向量。

(2) S对加法封闭，即若$S_i\in S$，$S_j\in S$，则$S_i+S_j\in S$。

如果(n, k)分组码的2^k个长度为n的码字构成了B_n的一个子空间，它就是线性码。因此，若C_i和C_j是(n, k)线性分组码的两个码字，那么C_i+C_j必定也是该码的码字。 231

例8.1：向量空间B_3包括所有长度为3的二进制向量：

$$B_3=\{[000], [001], [010], [011], [100], [101], [110], [111]\}$$

注意，B_3是它自身的一个子空间，因为它包含了全零矢量，并对加法封闭。判断B_3的下列子集中哪一个是子空间：

$A_1=\{[000], [001], [100], [101]\}$

$A_2=\{[000], [100], [110], [111]\}$

$A_3=\{[001], [100], [101]\}$

解：容易验证出A_1是子空间，因为它包含了全零向量，同时A_1中的任何两个向量的和都包含在A_1中。A_2不是子空间，它对加法不封闭，如$110+111=001\notin A_2$。A_3也不是子空间，因

为它既对加法不封闭：001+001=000$\notin A_3$，也不包含全零向量。

直观来看，如果一个码中码字之间的距离越大，则因信道中的错误致使发送码字被错译的机会也就越小。定义两个码字C_i和C_j之间的汉明距离（Hamming distance）为这两个码字之间不同元素的个数，记为$d(C_i, C_j)$或d_{ij}：

$$d_{ij} = \sum_{l=1}^{n}(C_i(l) + C_j(l)) \tag{8-1}$$

此处$C_m(l)$指C_m的第l个比特。例如对$C_i=[00101]$，$C_j=[10011]$，有$d_{ij}=3$。定义码字C_i中1的个数为其码重，因此$C_i=[00101]$的码重是2。C_i的码重就是它与全零码字$C_0=[00\cdots 0]$之间的汉明距d_{0i}，也就是该码字的所有元素之和：

$$w(C_i) = \sum_{l=1}^{n} C_i(l) \tag{8-2}$$

因为0+0=1+1=0，所以两个码字之间的汉明距就等于这两个码字之和的码重。例如对于$C_i=[00101]$，$C_j=[10011]$，有$w(C_i)=2$，$w(C_j)=3$，$d_{ij}=w(C_i+C_j)=w([10110])=3$。任意两个码字之间的汉明距等于其和的码重，而这个和也是一个码字，所以我们只需观察所有非零码与全零码C_0之间的最小距离，就能得出该码的最小汉明距：

$$d_{\min} = \min_{i, i\neq 0} d_{0i} \tag{8-3}$$

232　在8.2.6节中将会看到，线性分组码的最小码距是决定错误率的一个关键参数。

8.2.2　生成矩阵

生成矩阵反映如何从信息比特产生出线性分组码的码字。线性分组码的设计就是要找到这样一个生成矩阵，它所对应的码容易实现编解码，同时又有强大的纠错和检错能力。考虑（n, k）码的一个码字，其k个信息比特为：

$$U_i=[u_{i1},\cdots, u_{ik}]$$

编码后的码字为

$$C_i=[c_{i1},\cdots, c_{in}]$$

编码操作可用下面的n个方程来表示

$$c_{ij} = u_{i1}g_{1j} + u_{i2}g_{2j} + \cdots + u_{ik}g_{kj} \quad j = 1,\cdots,n \tag{8-4}$$

其中g_{ij}是二进制的0或1。这n个方程可写成矩阵形式

$$C_i=U_iG \tag{8-5}$$

其中的$k\times n$矩阵G就是这个码的生成矩阵（generator matrix），定义为

$$G = \begin{bmatrix} g_{11} & g_{12} & \cdots & g_{1n} \\ g_{21} & g_{22} & \cdots & g_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ g_{k1} & g_{k2} & \cdots & g_{kn} \end{bmatrix} \tag{8-6}$$

若记G中的第l行为$g_l=[g_{l1},\cdots, g_{ln}]$，那么任意一个码字$C_i$都可以表示成这些行向量的线

性组合：

$$C_i = u_{i1}g_1 + u_{i2}g_2 + \cdots + u_{ik}g_k \tag{8-7}$$

由于（n, k）线性分组码是n维空间上的一个k维子空间，因此G的k个行矢量$\{g_i\}_{i=1}^{k}$必须线性无关，使得它们能够张成这2^k个码字所对应的k维子空间。因此G的秩为k。子空间的基不唯一，所以生成矩阵也不唯一。

系统线性分组码的生成矩阵有如下形式：

$$G = [I_k \mid P] = \left[\begin{array}{cccc|cccc} 1 & 0 & \cdots & 0 & p_{11} & p_{12} & \cdots & p_{1(n-k)} \\ 0 & 1 & \cdots & 0 & p_{21} & p_{22} & \cdots & p_{2(n-k)} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & p_{k1} & p_{k2} & \cdots & p_{k(n-k)} \end{array}\right] \tag{8-8}$$

其中I_k是一个$k \times k$的单位阵，P是一个$k \times (n-k)$的矩阵，它决定检错和纠错所需的冗余比特或称校验位。系统码编码后的码字有如下形式：

$$C_i = U_iG = U_i[I_k \mid P] = [u_{i1}, \cdots, u_{ik}, p_1, \cdots, p_{(n-k)}] \tag{8-9}$$

其前k个比特是原始的信息位，后$n-k$个比特是根据信息位生成的校验位。按如下方式产生：

$$p_j = u_{i1}p_{1j} + \cdots + u_{ik}p_{kj}, \quad j = 1, \cdots, n-k \tag{8-10}$$

任何（n, k）线性分组码的生成矩阵都能通过行运算和列交换转化成系统生成矩阵。 233

例8.2：一般用连接到移位寄存器的$n-k$个模2加法器来实现系统线性分组码。将这样得到的校验位附在信息位后面就形成了码字。试画出下列生成矩阵G所对应的二进制（7, 4）码的实现结构。

$$G = \left[\begin{array}{cccc|ccc} 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 \end{array}\right] \tag{8-11}$$

解：矩阵G已经是系统形式，其P阵为

$$P = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \tag{8-12}$$

令P_{ij}表示矩阵P中第l行第j列的元素。由式（8-10）可知，码字中的第一个校验比特是$p_1 = u_{i1}P_{11} + u_{i2}P_{21} + u_{i3}P_{31} + u_{i4}P_{41} = u_{i1} + u_{i2}$。类似可得，第二个校验比特是$p_2 = u_{i1}P_{12} + u_{i2}P_{22} + u_{i3}P_{32} + u_{i4}P_{42} = u_{i1} + u_{i4}$，第三个校验比特是$p_3 = u_{i1}P_{13} + u_{i2}P_{23} + u_{i3}P_{33} + u_{i4}P_{43} = u_{i2} + u_{i3}$。图8-2是相应的移位寄存器实现结构。输出码字是$[u_{i1}u_{i2}u_{i3}u_{i4}p_1p_2p_3]$，开关置下时输出系统信息比特$u_{ij}(j = 1, \cdots, 4)$，开关置上时输出校验比特$p_j(j = 1, 2, 3)$。

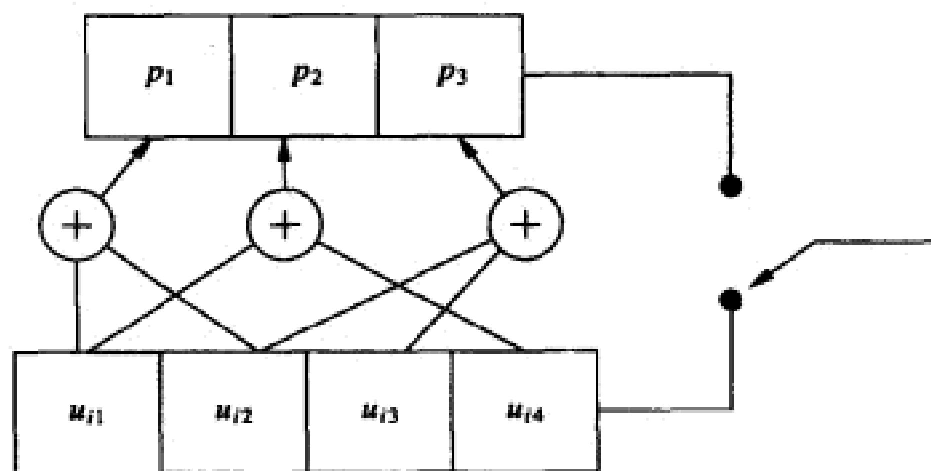


图8-2　二进制（7, 4）码的实现

8.2.3　校验矩阵和伴随式

校验矩阵用来译出生成矩阵为G的线性分组码。与生成矩阵$G=[I_k|P]$对应的校验矩阵H为

234

$$H=[P^T \mid I_{n-k}] \tag{8-13}$$

容易验证$GH=0_{k,n-k}$，其中$0_{k,n-k}$是一个$k\times(n-k)$的全零矩阵。因为码字C_i是信息比特U_i与生成矩阵G的乘积，即$C_i=U_iG$，所以对任意的输入序列U_i有

$$C_iH^T=U_iGH^T=0_{n-k} \tag{8-14}$$

其中0_{n-k}指长为$n-k$的全零行向量。因此，任意一个合法码字和校验矩阵的乘积是全零向量。可根据这个特性来判断接收向量是不是一个合法码字，这就是伴随式检测的思想。

令R表示对应发送码字C的接收码字。如果信道中没有出现差错，则$R=C$。当传输出现差错时，R中有一个或者多个码字符号将与C中的不同。因此，可将接收码字写成

$$R=C+e \tag{8-15}$$

其中的$e=[e_1, e_2,\cdots, e_n]$称为错误图样（error pattern），它表示通过信道后哪些符号发生了错误。定义R的伴随式（syndrome）为

$$S=RH^T \tag{8-16}$$

如果R是一个合法的码字（即存在某个i使$R=C_i$），那么由式（8-14）可知$S=C_iH^T=0_{n-k}$。所以，如果传输的码字没有发生错误，或者如果发生的错误图案正好使接收码字错成另一个合法码字，那么伴随式将是一个全零矢量。如果接收码字R中的错误是可检测的，那么$S\neq 0_{n-k}$。如果接收码字R中的错误是可纠正的，那么伴随式可以识别出所发生的错误图案，使这个错误被纠正。伴随式实际上只是错误图案e的函数，与传输码字C无关，因为

$$S=RH^T=(C+e)H^T=CH^T+eH^T=0_{n-k}+eH^T \tag{8-17}$$

$S=eH^T$对应于包含n个未知量的$n-k$个方程，所以对于给定的伴随式S，可解出2^k种可能的错误图案。一般情况下，比特出现错误的概率很小，且各比特是否出错是相互独立的随机事件，因此可能性最大的错误图案就是这2^k种可能图案中码重最小的那个，也就是比特错误个数最少的那个。因此，若对于给定的伴随式S，对应的最可能的错误图案样是$\hat{e}$，那么一般将发送码字译为

$$\hat{C}=R+\hat{e}=C+e+\hat{e} \tag{8-18}$$

如果实际发生的错误e就是最可能的错误图案$\hat{e}$，则$\hat{C}=C$，错误就被纠正了。具体的译码过程和相关的错误率将在8.2.6节讨论。

令C_w为（n, k）码中除全零码外码重最小的码字。由于$d_{\min}$是C_w中1的个数，所以C_wH^T是H的$d_{\min}$个列之和的转置。而H的秩至多是$n-k$，所以从$C_wH^T=0_{n-k}$可知（n, k）线性分组码的$d_{\min}$有如下上界

$$d_{\min}\leqslant n-k+1 \tag{8-19}$$

235

这个界叫Singelton界。

8.2.4 循环码

循环码是线性分组码的一个子集，其码字具有循环移位关系。比方说，如果$C=[c_0, c_1,\cdots, c_{n-1}]$是循环码的一个码字，则其一位循环移位$C^{(1)}$，即$[c_{n-1}, c_0,\cdots,c_{n-2}]$也是循环码的一个码字。更一般地，任意循环移位$C^{(i)}=[c_{n-i}, c_{n-i+1},\cdots,c_{n-i-1}]$都是循环码的一个码字。这种循环特性使得循环码的编解码更为简单，其复杂度大大低于一般线形分组码编码译码中的矩阵乘法运算。所以，实际中使用的大部分线性分组码都是循环码。

生成循环码所用的不是生成矩阵，而是*生成多项式*（generator polynomial）。（n, k）循环码的生成多项式$g(X)$是一个$n-k$次多项式：

$$g(X)=g_0+g_1X+\cdots+g_{n-k}X^{n-k} \tag{8-20}$$

式中g_i是二进制的0或1，$g_0=g_{n-k}=1$。将k比特的信息序列$[u_0,\cdots, u_{k-1}]$写成一个消息多项式就是：

$$u(X)=u_0+u_1X+\cdots+u_{k-1}X^{k-1} \tag{8-21}$$

与其对应的码字$C=[c_0,\cdots, c_{n-1}]$可通过生成多项式和消息多项式的乘积得到：

$$c(X)=u(X)g(X)=c_0+c_1X+\cdots+c_{n-1}X^{n-1} \tag{8-22}$$

当且仅当$c(X)$能被$g(X)$整除时，即

$$\frac{c(X)}{g(X)}=q(X) \tag{8-23}$$

时，$c(X)$所对应的码字是一个合法码字。上式中$q(X)$的次数小于k。

例8.3：*已知（7, 4）循环码的生成多项式是$g(X)=1+X^2+X^3$，试判断多项式$c_1(x)=1+X^2+X^5+X^6$和$c_2(X)=1+X^2+X^3+X^5+X^6$对应的码字是不是合法码字。*

236

解：*二进制多项式的除法与标准的多项式除法类似，差别只是二元域运算中，减法就是加法。将$c_1(X)$除以$g(X)$，我们有*

$$\begin{array}{r|l} & X^3+1 \\ \hline X^3+X^2+1 & X^6+X^5+X^2+1 \\ & \underline{X^6+X^5+X^3} \\ & \qquad\qquad X^3+X^2+1 \\ & \qquad\qquad \underline{X^3+X^2+1} \\ & \qquad\qquad 0. \end{array}$$

因为$g(X)$能整除$c(X)$，所以它是一个合法码字。实际上，$c_1(X)=(1+X^3)g(X)=u(X)g(X)$，因此与$c_1(X)$相应的信息比特序列是$U=[1001]$，它是消息多项式$u(X)=1+X^3$的系数。

将$c_2(X)$除以$g(X)$，我们有：

$$\begin{array}{r} X^3+1 \\ X^3+X^2+1 \overline{)\, X^6+X^5+X^3+X^2+1} \\ \underline{X^6+X^5+X^3 \qquad\quad} \\ X^2+1 \end{array}$$

结果有余项X^2+1，所以$c_2(X)$不是合法码字。

在线性分组码中我们曾经提到，系统码的前k个码字符号是信息比特，余下的是校验比特。循环码也可以编为系统码的形式，方法是先用X^{n-k}乘以消息多项式$u(X)$得到

$$X^{n-k}u(X)=u_0X^{n-k}+u_1X^{n-k+1}+\cdots+u^{k-1}X^{n-1} \tag{8-24}$$

这实际就是将信息比特置为码字多项式的高k位。再用$g(X)$去除式（8-24），得到

$$\frac{X^{n-k}u(X)}{g(X)}=q(X)+\frac{p(X)}{g(X)} \tag{8-25}$$

其中多项式$q(X)$的次数不超过$k-1$，余式$p(X)$的次数不超过$n-k-1$。用$g(X)$乘以式（8-25），得到

$$X^{n-k}u(X)=q(X)g(X)+p(X) \tag{8-26}$$

两边同时加上$p(X)$：

$$p(X)+X^{n-k}u(X)=q(X)g(X) \tag{8-27}$$

上式表明$p(X)+X^{n-k}u(X)$是一个合法的码字，因为它能被$g(X)$整除。该码字对应多项式$p(X)+X^{n-k}u(X)$的n个系数。注意$p(X)$的次数最大为$n-k-1$，可写成：

$$p(X)=p_0+p_1X+\cdots+p_{n-k-1}X^{n-k-1} \tag{8-28}$$

结合式（8-24）和式（8-28），我们有

$$\begin{aligned} & p(X)+X^{n-k}u(X) \\ & \quad = p_0+p_1X+\cdots+p_{n-k-1}X^{n-k-1}+u_0X^{n-k}+u_1X^{n-k+1}+\cdots+u_{k-1}X^{n-1} \end{aligned} \tag{8-29}$$

这也就是说，该码字的前k个比特是信息比特$[u_0,\cdots,u_k]$，其余$n-k$个比特是校验比特$[p_0,\cdots,p_{n-k-1}]$，说明它是系统码。

237

系统码的码字多项式是通过以下三步生成的：首先用X^{n-k}去乘消息多项式$u(X)$，然后将得到的结果$X^{n-k}u(X)$除以$g(X)$，得到余式$p(X)$（同时也能得到商式$q(X)$，但编码不需要这个商式），最后将$p(X)$加上$X^{n-k}u(X)$，就得到式（8-29）。多项式的乘法很容易实现，多项式除法可通过反馈移位寄存器（参考文献[1]的8.1节及文献[2]的6.7节）很容易地实现。因此系统循环码编码的成本和复杂度都非常低。

现在考虑怎样描述循环码中的差错。对应发送码字的码字多项式可写成：

$$c(X)=u(X)g(X) \tag{8-30}$$

接收码字也可写成多项式的形式：

$$r(X)=c(X)+e(X)=u(X)g(X)+e(X) \tag{8-31}$$

式中$e(X)$是错误多项式，系数为1的项表示对应位置发生了错误。例如发送码字是$\boldsymbol{C}=[1011001]$，接收码字是$\boldsymbol{R}=[1111000]$，那么错误多项式为$e(X)=X+X^{n-1}$。伴随式$s(X)$定义为$r(X)$除以$g(X)$得到的余式，其次数最高为$n-k-1$。由式（8-31）可知，伴随式$s(X)$等于错误多项式$e(X)$模$g(X)$。通过一个类似于编码时所用的那种移位寄存器除法电路可得到伴随式，如前所述，这个除法电路通常使用反馈移位寄存器来实现，因此其成本和复杂度都很低。

8.2.5 硬判决译码

线性分组码的错误概率与译码器是软判决还是硬判决有关。硬判决译码（hard decision decoding）中，每个发送的编码比特在接收端被解调为0或者1。例如在BPSK中，如果接收到的符号更靠近$\sqrt{E_b}$，就判为1，如果更靠近$-\sqrt{E_b}$，就判为0。这种解调方式可能会损失一些有用的信息。例如在BPSK中，如果译码器在判断发送码字的时候能够利用接收信号到$\sqrt{E_b}$和$-\sqrt{E_b}$之间的距离的话，可以得到比硬判决译码更好的性能。这种利用距离信息进行译码的方式就是软判决译码。无线通信系统中软判决译码要比硬判决译码更常见，我们将在8.2.7节中进行介绍。

硬判决译码采用基于汉明距离的最小距离译码。接收码字首先被解调为0或1，然后将解调器的输出送入译码器。译码器将收到的码字与所有2^k个可能的码字进行比较，然后选出离接收码字汉明距离最小，也即差别最少的那个作为译码结果。用数学语言来说就是

$$\text{选出}\boldsymbol{C}_j\text{使得}d(\boldsymbol{C}_j,\ \boldsymbol{R})\leqslant d(\boldsymbol{C}_i,\ \boldsymbol{R})\quad \forall i\neq j \tag{8-32}$$

如果与$\boldsymbol{R}$有最小距离的码字有多个，则随机选一个作为译码结果。

238

最大似然译码从所有合法码字中选出最有可能形成接收码字的一个作为译码结果。即若接收码字为$\boldsymbol{R}$，则最大似然译码器选择的码字为

$$\boldsymbol{C}_j=\arg\max_i\ p(\boldsymbol{R}\mid \boldsymbol{C}_i)\qquad i=0,\cdots,2^k-1 \tag{8-33}$$

在AWGN信道中，最可能出现的错误事件就是错误个数最少的事件，因此式（8-32）的最小距离判决准则和式（8-33）中的最大似然判决准则是等价的。译码器得到最大似然码字$\boldsymbol{C}_i$后，再将它译为产生码字$\boldsymbol{C}_i$的k个信息比特。

最大似然译码是基于距离度量进行译码的，图8-3所示的码空间可以清楚地说明译码过程。图中的黑点表示码字，阴影点表示存在比特错误的接收码字。码字之间的最小汉明距离是$d_{\min}$。每个码字位于一个半径为$t=\lfloor 0.5d_{\min}\rfloor$的球体的球心，$\lfloor x\rfloor$代表向下取整。图中还标出了$\boldsymbol{C}_1$与$\boldsymbol{C}_2$之间的汉明距离。

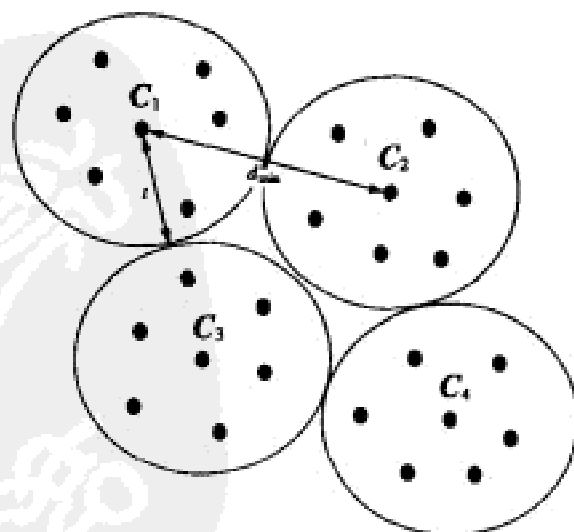


图8-3 码空间上的最大似然译码

最小距离译码除了可用于纠错外，也可用于检错。译码器发现数据分组中有错时，不对

接收数据做处理，可能直接丢弃，也可能等待重传。如果是纠错的话，译码器将把接收数据中的错误比特进行反转，使错误比特被纠正。对于纠错，如果接收码字离发送码字的汉明距离不超过t，最小距离译码就能保证译对，即可纠正多达t个错误。在图8-3中，错误个数不超过t的接收码字将会落在以发送码字为圆心的圆内，最小距离译码自然会译到圆心的那个码字。对于检错，如果错误数不超过$d_{min}-1$个，则一定可以检出。(n, k)码实际可以检出2^n-2^k个可能的错误图案。因为总共有2^n-1个不同的错误图案，其中2^k-1个错误图案将使发送码字错为另一个不同的合法码字（一共有2^k个合法码字），其余的都是可检出的。注意这不是硬判决译码，因为只有检错没有纠错。

239

例8.4：假设（5，2）码的码字为$C_0=[00000]$、$C_1=[01011]$、$C_2=[10101]$、$C_3=[11110]$，今传送全零码字C_0，请找出所有不可检出的错误图案。

解：不可检的错误图案对应于题中的三个非零码字，即为$e_1=[01011]$、$e_2=[10101]$、$e_3=[11110]$。这些错误图案与C_0相加后依然是合法码字。

8.2.6 AWGN信道中的硬判决译码的错误率

码字错误率P_e定义为发送码字被译错的概率。硬判决译码时，如果接收码字中的错误数超过t个的话可能会被译错。但注意，也有可能出现接收码字中的错误数超过了t，但它仍然离发送的码字最近，此时不会发生译码错误。因此，译码发生错误的概率小于等于错误个数超过t的概率。AWGN信道中各个比特发生错误的可能性是独立的，故有

$$P_e \leqslant \sum_{j=t+1}^{n} \binom{n}{j} p^j (1-p)^{n-j} \tag{8-34}$$

其中p是比特传输的错误概率，也即第6章所讲的无编码时的误比特率。例如对于BPSK有$p=Q\left(\sqrt{2E_c/N_0}\right)$，其中$E_c$是编码后每个符号的能量，$N_0/2$是噪声的功率谱密度。每个码字符号包含$k/n$个信息比特，因此每信息比特能量和每码字符号能量之间的关系为$E_c=kE_b/n$。增加分组码的校验比特数将增强纠错能力，但也降低了每个码字的符号能量，这样一来，码字符号的错误率p将增加。对于中高信噪比，编码的纠错能力可以超过p的增加，从而能产生出正的编码增益。但在低信噪比时，有可能出现负的编码增益，使得有编码的误码率甚至高于无编码的误码率。如果一个译码器能纠正发生在一个码字中的t个或者更少的错误，而不能纠正多于t个的错误，则这样的码称为完备码。对于完备码，式（8-34）中等式成立。

240

高信噪比时最可能发生的码字错误是错判为发送码字的最近邻码字。从最近邻差错可以导出译码错误率的上界和下界。下界是发送码字错为距离d_{min}处的特定码字概率：

$$P_e \geqslant \sum_{j=t+1}^{d_{min}} \binom{d_{min}}{j} p^j (1-p)^{d_{min}-j} \tag{8-35}$$

上界是一个联合界。假设所有其他2^k-1个码字与发送码字的汉明距都是d_{min}，误码率的联合界就是2^k-1乘上错为其中一个的概率[式（8-35）中的概率]：

$$P_e \leqslant (2^k-1)\sum_{j=t+1}^{d_{\min}}\binom{d_{\min}}{j}p^j(1-p)^{d_{\min}-j} \tag{8-36}$$

当码字数目很大或信噪比很低时，上面这两个界比较松。

用切尔诺夫界可得到码字错码概率更紧的界。对于标准正态分布的随机变量X，切尔诺夫界指出$P(X>x)\leqslant e^{-x^2/2}$。文献[3]的5.2节利用切尔诺夫界得出全零码字被错译为第j个码字的概率的上界是

$$P(w_j)\leqslant[4p(1-p)]^{w_j/2} \tag{8-37}$$

其中w_j是第j个码字的码重。全零码可能被译错为其他2^k-1个合法码字中的任一个，因此码字错误率的上界为

$$P_e \leqslant \sum_{j=1}^{2^k-1}[4p(1-p)]^{w_j/2} \tag{8-38}$$

该界需要所有码字的码重分布$\{w_j\}_{j=1}^{2^k-1}$（$j=0$对应全零码字）。将式（8-38）中的码重w_j换为$d_{\min}$可得到一个更简单，但略松的上界

$$P_e \leqslant (2^k-1)[4p(1-p)]^{d_{\min}/2} \tag{8-39}$$

注意码字错误率P_e与p有关，而p又与发送码字符号时所用的调制星座点之间的欧氏距离有关。因此，AWGN信道中好的编码设计不应当基于汉明距离，而应该使调制后码字之间的欧式距离最大化。这就要求信道编码和调制进行联合设计，8.7节将要介绍的编码调制就是这样一种思路。不过在8.8节中我们将看到，编码和交织结合使用时，汉明距离更能反映衰落情形下编码的性能。

241

译码后的比特错误率与特定的码及译码器有关，特别是与比特到码字的映射关系有关。比特到码字的映射如同多进制调制中比特到星座点的映射。经常将译码后的比特错误率近似为（见文献[2]的6.5节）：

$$P_b \approx \frac{1}{n}\sum_{j=t+1}^{n} j\binom{n}{j}p^j(1-p)^{n-j} \tag{8-40}$$

对于$t=1$的情形，上式简化为[2]$P_b\approx p-p(1-p)^{n-1}$。

例8.5：我们在8.2.8节将要提到的戈雷码（Golay code）可扩展为一个（24, 12）线性分组码，其最小码距为$d_{\min}=8$。假设码字在AWGN信道中以BPSK调制传输，$E_b/N_0=10$dB。请用式（8-39）的界计算P_e，再用近似式$P_b=P_e/k$计算误比特率P_b，并同无编码的误比特率进行比较。

解：由$E_b/N_0=10$dB得$E_c/N_0=\frac{12}{24}\times10=5$。因此$p=Q(\sqrt{10})=7.82\times10^{-4}$。将$k=12$和$d_{\min}=8$代入式（8-39）得$P_e\leqslant3.92\times10^{-7}$。再用误比特率的近似式得到$P_b\approx(1/k)P_e=3.27\times10^{-8}$。无编码时的误比特率为$P_b=Q\left(\sqrt{2E_b/N_0}\right)=Q\left(\sqrt{20}\right)=3.87\times10^{-6}$。可见这种编码使性能提高了两个数量级。注意式（3-89）是松散界，它和真实错误率可能会相差好几个数量级，因此本例中的计算结果有可能大大低估了该码的性能。

8.2.7 AWGN信道中软判决译码的错误率

硬判决译码丢弃了一些对译码有用的信息。例如对于BPSK，发送信号的星座点是$\pm\sqrt{E_b}$，接收端对匹配滤波后的接收值进行硬判决，即若接收值离$\sqrt{E_b}$更近时判为1，离$-\sqrt{E_b}$更近时判为0。译码器用这个硬判决结果进行译码，没有用到接收值离$\sqrt{E_b}$和$-\sqrt{E_b}$的具体距离。如果译码器使用了这些距离，就称为**软判决译码**（soft decision decoding，SDD）。在软判决译码中，解调器不对发送的值是0还是1做出硬性判断，而是输出一个软值，这个值对应于接收到的符号和发送符号（0或1）之间的距离。下面我们讨论BPSK调制下软判决译码的基本思路，这些思路很容易推广到高阶调制中。

设码字经过BPSK调制后在信道中传输。与硬判决一样，每个码字符号的能量为$E_c=(k/n)E_b$。对应第j个码字符号是1或者0，接收值分别是$r_j=\sqrt{E_c}+n_j$和$r_j=-\sqrt{E_c}+n_j$，其中n_j是均值为0、方差为$N_0/2$的加性白高斯噪声。在软判决译码中，对于给定的接收码字$\boldsymbol{R}=[r_1,\cdots,r_n]$，译码器对每个合法码字$\boldsymbol{C}_i(i=0,\cdots,2^k-1)$计算出一个相关度量$C(\boldsymbol{R},\boldsymbol{C}_i)$，然后选出相关度量最大的$\boldsymbol{C}_i$作为译码结果。相关度量定义为

242

$$C(\boldsymbol{R},\boldsymbol{C}_i)=\sum_{j=1}^{n}(2c_{ij}-1)r_j \tag{8-41}$$

式中c_{ij}是码字$\boldsymbol{C}_i$的第j个比特。对应c_{ij}取值为1、0，$2c_{ij}-1$取值为1和−1。所以相关度量就是用比特传输时的极性对接收值做加权和。当接收码字符号中的多数与$\boldsymbol{C}_i$对应的发送符号同极性且值比较大时，相关度量$C(\boldsymbol{R},\boldsymbol{C}_i)$比较大；当大部分接收到的符号值很小，并与$\boldsymbol{C}_i$对应的符号同极性时，相关度量比较小；如果大部分接收到的符号与$\boldsymbol{C}_i$对应的发送符号反极性时，相关度量一般是负值。当信噪比非常高时，如果发送的是码字$\boldsymbol{C}_i$，则有$C(\boldsymbol{R},\boldsymbol{C}_i)\approx n\sqrt{E_c}$，当$j\neq i$时，$C(\boldsymbol{R},\boldsymbol{C}_j)<n\sqrt{E_c}$。

在AWGN信道中，线性码的任何码字有相同的错误率。因此，为了简化错误率分析，一般都假设发送的是全零码字。我们也假设发送全零码字$\boldsymbol{C}_0$，对应的接收码字为$\boldsymbol{R}$。如果$C(\boldsymbol{R},\boldsymbol{C}_0)>C(\boldsymbol{R},\boldsymbol{C}_i)$，$i=1,\cdots,2^k-1$，译码就是正确的。记第$i$个码字$\boldsymbol{C}_i$的码重为$w_i$（$w_i$等于$\boldsymbol{C}_i$中1的个数），发送码字$\boldsymbol{C}_i$的条件下，$C(\boldsymbol{R},\boldsymbol{C}_i)$服从均值为$\sqrt{E_c}n(1-2w_i/n)$、方差为$nN_0/2$的高斯分布。注意各个相关度量之间不是相互独立的，它们都和同一个接收码字$\boldsymbol{R}$有关。可以证明，概率$P_e(\boldsymbol{C}_i)=p(C(\boldsymbol{R},\boldsymbol{C}_0)<C(\boldsymbol{R},\boldsymbol{C}_i))$就是方差为$2w_iN_0$的高斯随机变量小于$-2w_i\sqrt{E_c}$的概率，即

$$P_e(\boldsymbol{C}_i)=Q\left(\frac{2w_i\sqrt{E_c}}{\sqrt{2w_iN_0}}\right)=Q\left(\sqrt{2w_i\gamma_bR_c}\right) \tag{8-42}$$

根据联合界，对每个$\boldsymbol{C}_i$的成对错误率求和可得到错误概率的上界为

$$P_e\leqslant\sum_{i=1}^{2^k-1}P_e(\boldsymbol{C}_i)=\sum_{i=1}^{2^k-1}Q\left(\sqrt{2w_i\gamma_bR_c}\right) \tag{8-43}$$

这个界需要已知码重分布$w_i(i=1,\cdots,2^k-1)$。注意到$w_i\geqslant d_{\min}$，所以上式可以简化为

$$P_e\leqslant(2^k-1)Q\left(\sqrt{2\gamma_bR_cd_{\min}}\right) \tag{8-44}$$

利用Q函数的切尔诺夫界 $Q(\sqrt{2x})<e^{-x}$ 可进一步得到

$$P_e \leqslant (2^k-1)e^{-\gamma_b R_c d_{\min}} < 2^k e^{-\gamma_b R_c d_{\min}} = e^{-\gamma_b R_c d_{\min}+k\ln 2} \tag{8-45}$$

而未编码的BPSK的误比特率为

$$P_b = Q(\sqrt{2\gamma_b}) < e^{-\gamma_b} \tag{8-46}$$

将上面两式相比较可知分贝值的编码增益大约为

$$G_c = 10\log_{10}[(\gamma_b R_c d_{\min} - k\ln 2)/\gamma_b] = 10\log_{10}[R_c d_{\min} - (k\ln 2)/\gamma_b] \tag{8-47}$$

243

注意编码增益与编码率、码字中信息比特的个数、最小码距、信噪比等有关。编码增益随γ_b的降低而不断减小，信噪比小到一定程度时，编码增益成为负值。一般来说，软判决译码的性能要比硬判决译码好2dB~3dB（见文献[1]的8.1节）。

例8.6：对于例8.5中$d_{\min}=8$的（24,12）码，试估计$\gamma_b=10$dB时编码增益。

解：将$\gamma_b=10$、$R_c=12/24$、$d_{\min}=8$及$k=12$代入（8-47）得$G_c=5$dB。由于该码的最小码距很大，所以编码增益也很大。

8.2.8 常见的线性分组码

接下来我们介绍一些常见的线性分组码，更详细的内容请参考文献[1~4]。最常见的分组码是汉明码，它有一个参数$m\geqslant 2$。(n, k) 汉明码的码长是$n=2^m-1$、信息比特数是$k=2^m-m-1$、校验比特数为$n-k=m$。汉明码的最小码距是$d_{\min}=3$，可纠正发生在$n=2^m-1$个码字符号上的$t=1$个错误。虽然汉明码的纠错能力不很强，但因为它是完备码，所以其错误率满足式（8-34）中的等式。

戈雷码和扩展戈雷码是一种纠错能力很强的码。戈雷码是一个（23, 12）的线性分组码，其$d_{\min}=7$、$t=3$。给戈雷码增加一个偶校验比特就是扩展戈雷码，它是一个（24, 12）分组码，其$d_{\min}=8$、$t=3$。由于t未变，所以扩展的这个校验比特并没有增强纠错能力，但它使信息比特率正好等于传输比特率的一半，因此编码器的输入输出可以共用一个时钟，输入时钟是输出的二分频，这使得实现非常方便。戈雷码和扩展戈雷码的最小码距比汉明码更大，因此有更好的纠错能力，但代价是译码更复杂，同时码率也更低（$R_c=k/n$）。低码率意味着数据速率降低或者传输带宽增加。

另一种强纠错能力的分组码是BCH（Bose-Chadhuri-Hocquenghem）码。它是一种循环码。对于高编码率，BCH的性能在中高信噪比时一般比相同n、k的其他分组码更好。BCH码在码长、码率和纠错能力方面有很大的选择余地。常用BCH码的码长满足$n=2^m-1$，其中整数$m\geqslant 3$。

图8-4示出了一些BCH码在BPSK调制及硬判决译码下的误比特率曲线。图中的P_b是式（8-40）给出的近似值，其中BPSK调制的误比特率p为

$$p = Q\left(\sqrt{\frac{2E_c}{N_0}}\right) = Q(\sqrt{2R_c\gamma_b}) \tag{8-48}$$

从图中可见，BCH(127, 36)在信噪比很低时呈现为负的编码增益。这种现象在强纠错能

力的编码中并不少见。其原因在8.2.6节已指出，就是编码能力越强，每编码符号的能量就越小。

244

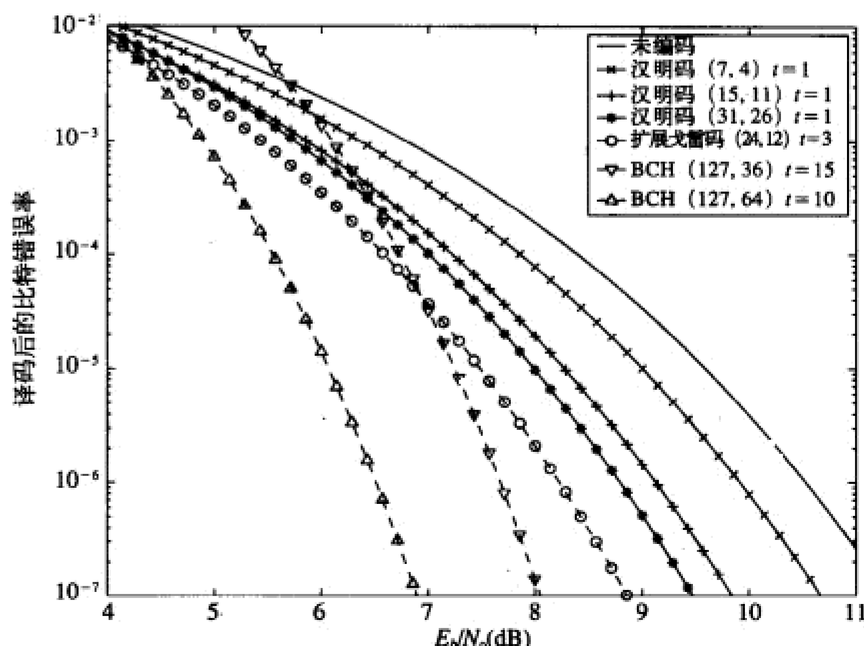


图8-4　不同BCH码的比特错误率

8.2.9　多进制分组码：Reed Solomon码

多进制分组码与二进制码类似：它将K个信息符号映射为N个编码符号，只不过符号是q进制的，其取值范围是$\{0,1,\cdots,q-1\}$。通常$q=2^k$，即一个符号对应k个比特。

最常见的多进制分组码就是Reed Soloman（RS）码，广泛应用于磁带录音、蜂窝数字分组数据（CDPD）等许多系统中。RS码的码长$N=q-1=2^k-1$，信息符号个数可以是$K=1, 2,\cdots, N-1$，不同的K值有不同的纠错能力：RS码可以纠正$t=0.5\lfloor N-K\rfloor$个符号错误。多进制码的码距定义为码字之间不相同的符号的数目。RS码的最小码距是$d_{\min}=N-K+1$，这是给定编码长度和输入符号数时，线性分组码最大可能的最小码距。RS码达到了这个最大可能的最小距离$d_{\min}$。实际应用中经常将RS码缩短以满足特定系统的要求（见文献[4]的5.10节）。

多进制码，特别是RS码的每个符号都对应k个比特，因此时常用于$M=2^k$的多进制调制中。采用2^k进制调制时，每个码字符号作为一个星座点传输。若相应的调制将发送星座点误判为其他星座点的概率为P_M（即误码率），那么多进制码的码字错误率上界为

$$P_e \leqslant \sum_{j=t+1}^{N}\binom{N}{j}P_M^j(1-P_M)^{N-j} \tag{8-49}$$

它在形式上与二进制码的对应公式（8-34）相似。信息符号的错误率可近似为

$$P_s \approx \frac{1}{N}\sum_{j=t+1}^{N} j\binom{N}{j} P_M^j (1-P_M)^{N-j} \tag{8-50}$$

245

8.3 卷积码

在卷积码中，信息比特通过图8-5所示的有限状态移位寄存器产生出编码符号。移位寄存器共有K个寄存器，每个寄存器存储k个比特。编码器有n个二进制的加法器，其输入取自K个寄存器。二进制输入比特通过移位寄存器移位，每次移位移入k个比特，对应产生n个编码比特。卷积码的编码率为$R_c = k/n$。每个输入比特最大能够影响到的输出符号的时间范围称为*约束长度*（constraint length）。在分组码中，每n比特输出只受输入的k比特影响。但对于卷积码，从图8-5中可以看到，每n比特输出要受到kK个输入比特的影响，或者说是K个k比特的字节。卷积码当前输出的n比特码字不仅与当前输入的k个信息比特有关，还和更多的输入比特有关（共与kK个比特有关），因此称卷积码是有记忆的编码。注意卷积码中每级寄存器存储的比特数也有可能不相同。

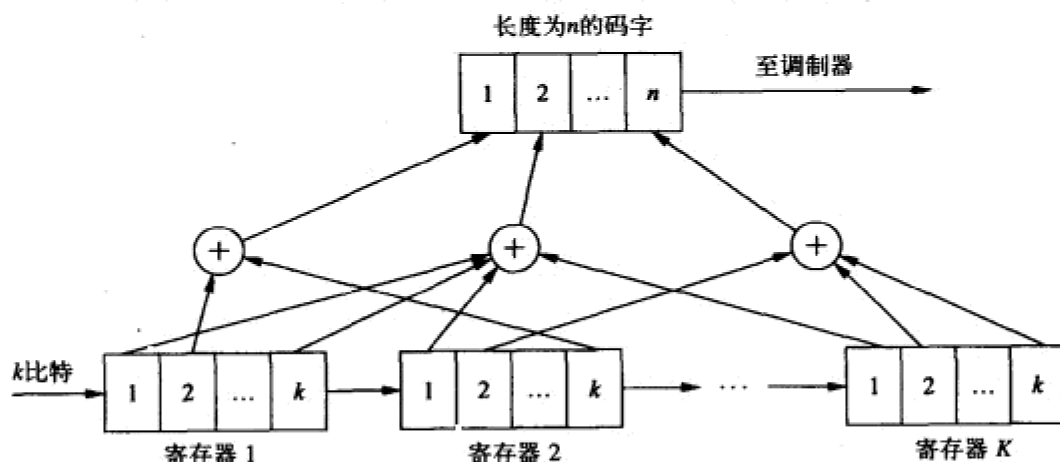


图8-5 卷积码编码器

8.3.1 卷积码的描述：格图

图8-5中的卷积码所产生的n比特码字不仅取决于第一个寄存器的k比特输入，还与其他$K-1$个寄存器的内容有关。其他$K-1$个寄存器的内容称为编码器的状态（state），它有$2^{k(K-1)}$个可能的取值。说明一个卷积码就需要说明输出码字和k个输入比特及编码器状态的关系。有许多描述卷积码的方法，包括树图、状态图和格图（见文献[1]的8.2节）。树图以树的形式表示编码器，每个树枝代表不同的编码器状态和对应的编码输出。状态图示出了编码中不同状态之间的转移关系及相应的编码输出。格图是树图的一种简化。当树深超过约束长度时，树图开始出现重复。将树图中相同状态的节点合为一个，就形成了格图。所以本节只讨论格图，它是描述卷积码最常用的方法。下面我们通过一个实例来说明。

246

考虑图8-6中所示的卷积码编码器，其中$n=3$、$k=1$、$K=3$。每个时刻有一个比特移入该三阶移位寄存器的第一级。编码器用这三个寄存器的内容来产生3bit的输出码字$C_1C_2C_3$。记t时刻第i个寄存器中的比特为S_i，那么从图中可以看出$C_1=S_1$、$C_2=S_1+S_2+S_3$、$C_3=S_1+S_3$。比特序列$\boldsymbol{U}$输入到编码器产生出编码符号序列$\boldsymbol{C}$。注意与C_1相对应的编码符号其实就是原始信息比特。当卷积码的编码符号中有一个就是原始的信息比特时，就说这个卷积码是系统码。定义编码器的状态$\boldsymbol{S}$为后两个寄存器的内容$\boldsymbol{S}=S_2S_3$，总共有$2^2=4$个可能的状态值。描述这个编码器就是要描述在不同的输入比特和状态下，编码器对应的输出以及下一个时刻状态的变化。

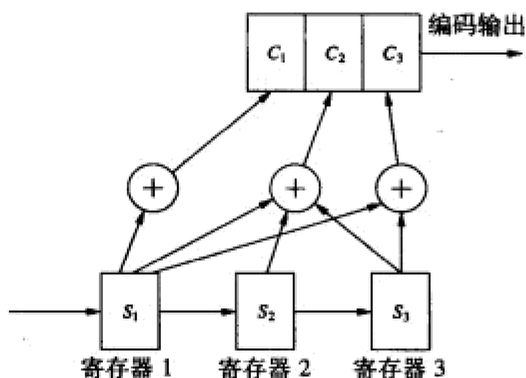


图8-6　卷积码编码器示例
($n=3$、$k=1$、$K=3$)

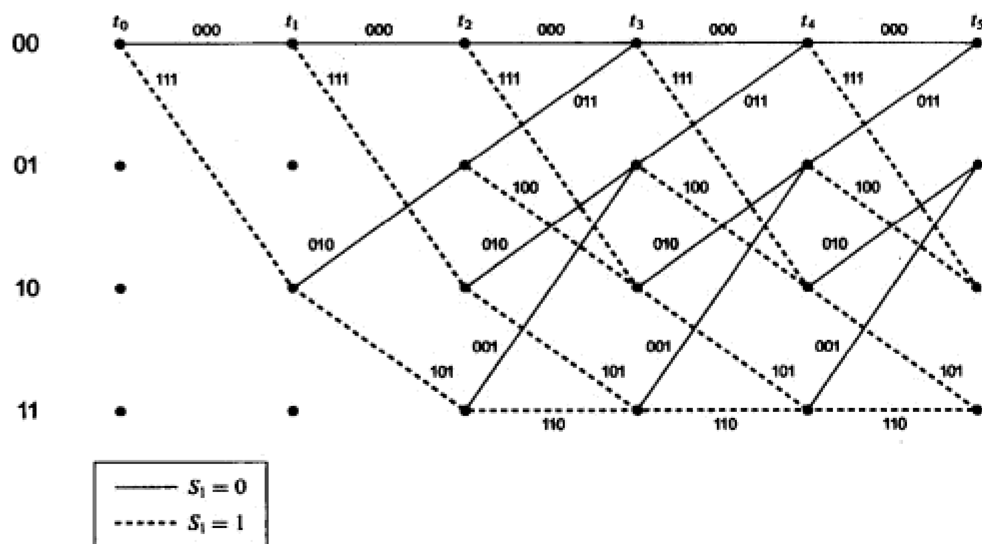


图8-7　格图

247

图8-7就是上面这个卷积码的格图。图中的实线代表输入比特为0时的状态转移，虚线表示输入比特为1时的状态转移。比如编码器的初始状态为$\boldsymbol{S}=00$，如果输入0到寄存器1，则在移存器移位之后，寄存器1中的0被移到了寄存器2中，寄存器2中的0又被移到了寄存器3中，于是编码器状态还是$\boldsymbol{S}=00$。如果输入1到寄存器1，则在移存器移位之后，寄存器1中的1被移到了寄存器2中，寄存器2中的0被移到了寄存器3中，新的状态将会变为$\boldsymbol{S}=10$。图8-7中状态转移线所标注的是状态为$\boldsymbol{S}$和输入比特为S_1时的编码输出，它由S_1、S_2和S_3通过加法操作得到。例如对于$\boldsymbol{S}=00$，$S_1=1$，编码输出$C_1C_2C_3$就是

$C_1=S_1=1$、$C_2=S_1+S_2+S_3=1$、$C_3=S_1+S_3=1$。图8-7中，从状态$S=00$转移到状态$S=10$的虚线标出了输出值111。注意$S_1=0$和$S=00$对应的编码输出总是一个全零码字，因为任意个0相加结果总是0。格图中时间t_i和t_{i+1}之间的部分称为格图的第i个分支。在t_0时刻，格图的初始状态是全零状态。在t_3时刻，格图达到稳态，所谓稳态是指格图中所有的状态都可以从前面的两个状态到达。到达稳态后，格图开始重复先前的过程。稳态时，从格图的每一个状态出发可到达下一时刻的两个状态。一般情况下，格图在t_0时刻从全零状态开始，t_K时刻达到稳态。

对于一般的k、K值，格图有2^{K-1}个状态，每个状态有2^k条路径离开，有2^k条路径到达。整个格图中的全部路径数目随k、K以及路径长度呈指数增长。

248

例8.7：考虑图8-7中的格图所表示的卷积码。假定初始状态为$S=S_2S_3=01$，试求输入比特序列为$U=011$时的状态序列S以及编码器输出C。

解：格图中在t_2时刻首次出现状态$S=01$。t_2时刻输入信息比特为$S_1=0$时，格图从t_2时刻的状态$S=01$出发，经过实线到达t_3时刻的状态$S=00$，这条路径对应的编码输出为$C=011$。再从t_3时刻的状态$S=00$出发，因为输入的信息比特为$S_1=1$，所以沿着虚线到达t_4时刻的状态$S=10$，对应的输出是$C=111$。最后，从t_4时刻的状态$S=10$出发，因为输入信息比特为$S_1=1$，所以沿着虚线到达t_5时刻的状态$S=11$，对应的编码器输出为$C=101$。

8.3.2 最大似然译码

(n, k)线性分组码的最大似然译码就是在所有合法码字中找出一个最接近接收码字的码字。与此不同，由有限状态移位寄存器产生的卷积码实质上是一个有限状态机，它的最大似然译码法则是：对于给定的接收符号序列R，找出最大可能的编码符号序列C。即译码输出的符号序列C^*满足

$$p(R \mid C^*) \geqslant p(R \mid C) \quad \forall C \tag{8-51}$$

每一个可能的序列C都对应格图上的一条路径，最大似然译码也就是寻找格图中最大可能的路径。AWGN信道中噪声对每个编码符号的影响是独立的，对于编码率为$1/n$的卷积码，格图中长度为L的路径的似然函数（8-51）可表示为

$$p(R \mid C)=\prod_{i=0}^{L-1} p(R_i \mid C_i)=\prod_{i=0}^{L-1}\prod_{j=1}^{n} p(R_{ij} \mid C_{ij}) \tag{8-52}$$

式中C_i和R_i分别是编码序列C及接收序列R的一部分，对应于格图的第i个分支。C_{ij}是C_i中的第j个编码符号，R_{ij}是R_i中的第j个接收符号。$p(R \mid C)$的对数值叫做对数似然函数：

$$\log p(R \mid C)=\sum_{i=0}^{L-1}\log p(R_i \mid C_i)=\sum_{i=0}^{L-1}\sum_{j=1}^{n}\log p(R_{ij} \mid C_{ij}) \tag{8-53}$$

其中对应于i支路的部分

$$B_i=\sum_{j=1}^{n}\log p(R_{ij} \mid C_{ij}) \tag{8-54}$$

249

叫做**分支度量**（branch metric）。由于对数是单调增函数，所以能使似然函数最大的路径必然也使对数似然函数最大。对数似然函数在运算上显然便利的多，它只有加减运算，没有乘法运算。格图中的一条路径所对应的对数似然函数称为**路径度量**（path metric），由式（8-53）可见，路径度量等于该路径上所有分支度量的和。具有最大路径度量的路径就是最大似然路径。

卷积码的译码可以是软判决译码也可以是硬判决译码。硬判决译码中R_{ij}是1或者0。令p代表解调的判决错误率，若$\boldsymbol{R}$和$\boldsymbol{C}$的长度为N，它们有d处不同（即汉明距为d），则有

$$p(\boldsymbol{R} \mid \boldsymbol{C}) = pd(1-p)^{N-d}$$

$$\log p(\boldsymbol{R} \mid \boldsymbol{C}) = -d\log\frac{1-p}{p} + N\log(1-p) \tag{8-55}$$

因为p小于0.5，所以式（8-55）的值在d最小时达到最大。因此，最大似然序列就是所有可能的编码序列中离$\boldsymbol{R}$的汉明距离最小的序列。

软判决译码中R_{ij}直接用接收值，而不是0、1这种判决值。比如对于BPSK，若发1映射为$\sqrt{E_c}$、发0映射为$-\sqrt{E_c}$，那么发送C_{ij}时有

$$R_{ij} = \sqrt{E_c}(2C_{ij} - 1) + n_{ij} \tag{8-56}$$

其中$E_c = kE_b/n$是每编码符号的能量，n_{ij}的是均值为零、方差为$\sigma^2 = N_0/2$的高斯噪声。因此：

$$p(R_{ij} \mid C_{ij}) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(R_{ij} - \sqrt{E_c}(2C_{ij} - 1))^2}{2\sigma^2}\right] \tag{8-57}$$

最佳译码就是在选择与R_{ij}欧氏距离最近的C_{ij}。对于两个不同的路径$\boldsymbol{C}_1$、$\boldsymbol{C}_2$，忽略式（8-53）中相同的项不会改变路径度量的大小次序。同样地，给式（8-53）中所有项乘一个系数也不会改变路径度量的大小次序。因此，忽略式（8-57）中的系数和公共项后，式（8-53）中的$\sum_{j=1}^{n}\log p(R_{ij} \mid C_{ij})$可以替换为下面的**等效分支度量**（equivalent branch metric）而不改变译码结果：

250

$$\mu_i = \sum_{j=1}^{n} R_{ij}(2C_{ij} - 1) \tag{8-58}$$

下面我们将对图8-6中的卷积码结合图8-7的格图说明路径度量的计算。为简单起见，只考虑格图中的两条路径，我们计算这两个路径在给定接收序列$\boldsymbol{R}$时的似然值。假设格图从t_0时刻的全零状态开始。第一条路径是全零路径，它对应于全零的输入序列。第二条路径是从t_0时刻的$S = 00$状态开始，t_1时刻到达状态$S = 10$，t_2时刻到达状态$S = 01$，最后于t_3时刻到达状态$S = 00$，与第一条路径重合。这两条路径在$t < t_0$及$t \geqslant t_3$时的分支度量都相同，所以最大似然路径就是不相同的这部分当中分支度量之和更小的那个。最大似然路径对应于不相同部分的分支度量之和比较小的路径。从图8-7可知，在前三个支路，全零路径产生的码字序列为$\boldsymbol{C}_0 = 000000000$，第二条路径产生的码字序列为$\boldsymbol{C}_1 = 111010011$。

先考虑错误率为p的硬判决译码。假定前三个支路对应的接收序列为$\boldsymbol{R} = 100110111$。

$\boldsymbol{R}$和C_0之间的汉明距为6，$\boldsymbol{R}$和C_1之间的汉明距为4。如前所述，最有可能的路径就是对应C_1的那条路径，因为C_1与$\boldsymbol{R}$之间的汉明距最小。全零路径的度量为：

$$M_0=\sum_{i=0}^{2}\sum_{j=1}^{3}\log P(R_{ij}\mid C_{ij})=6\log p+3\log(1-p) \tag{8-59}$$

另一条路径的度量为：

$$M_1=\sum_{i=0}^{2}\sum_{j=1}^{3}\log P(R_{ij}\mid C_{ij})=4\log p+5\log(1-p) \tag{8-60}$$

通常有$p<<1$，于是$M_0\approx 6\log p$、$M_1\approx 4\log p$。因为$\log p<0$，所以第二条路径的度量比第一条路径更大。

再考虑软判决译码。假设$E_c=1$，前三条支路对应的接收序列为$\boldsymbol{R}=(0.8, -0.35, -0.15, 1.35, 1.22, -0.62, 0.87, 1.08, 0.91)$，则全零路径的度量是：

$$M_0=\sum_{i=0}^{2}\mu_i=\sum_{i=0}^{2}\sum_{j=1}^{3}R_{ij}(2C_{ij}-1)=\sum_{i=0}^{2}\sum_{j=1}^{3}-R_{ij}=-5.11$$

第二条路径的度量为：

$$M_1=\sum_{i=0}^{2}\sum_{j=1}^{3}R_{ij}(2C_{ij}-1)=1.91$$

第二条路径的度量比第一条路径大。要判断第二条路径是否为所有路径中的最大似然路径，还需要将它与格图中的所有其他路径进行比较。

251

最大似然译码中，对数似然函数式（8-53）的计算必须要遍历格图中的所有路径，其计算复杂度随编码的存储量呈指数增加。下一节将要介绍的维特比算法充分利用了路径度量的结构特点，简化了最大似然译码的复杂度。

8.3.3 维特比算法

维特比于1967年提出的[5]维特比算法能够系统地把那些不可能具有最大度量的路径排除，从而降低了最大似然译码的复杂度。对于格图中的某个节点N，每一条到达它的路径在离开后所走的可能路径都是一样的。如果格图中的某个路径经过了节点N，并且具有最大路径度量，那么只看到达节点N之前的这些局部路径的话，这条路径在这些局部路径中也必然有最大的度量。参见图8-8，假设有三条路径在n时刻进入节点N，到达此节点时的累积路径度量为$P^l=\sum_{k=0}^{n-1}B_k^l (l=1,2,3)$，其中$P^1$最大。再假设全局度量最大的路径在$n$时刻经过了节点$N$，且该路径在节点$N$之后的分支度量为$\{B_k\}$，那么它在节点$N$之后的累积路径度量为$\sum_{k=n}^{\infty}B_k$。这样，节点$N$之前的三条路径和节点$N$之后的最大似然路径合起来形成的三条全局路径的度量为$P^l+\sum_{k=n}^{\infty}B_k, l=1,2,3$。由此可见，如果一条路径在到达节点$N$时有最大的局部路径度量，那么在所有经过节点$N$的全局路径中，它

也具有最大度量。前面例子中的路径1就是这样的例子。

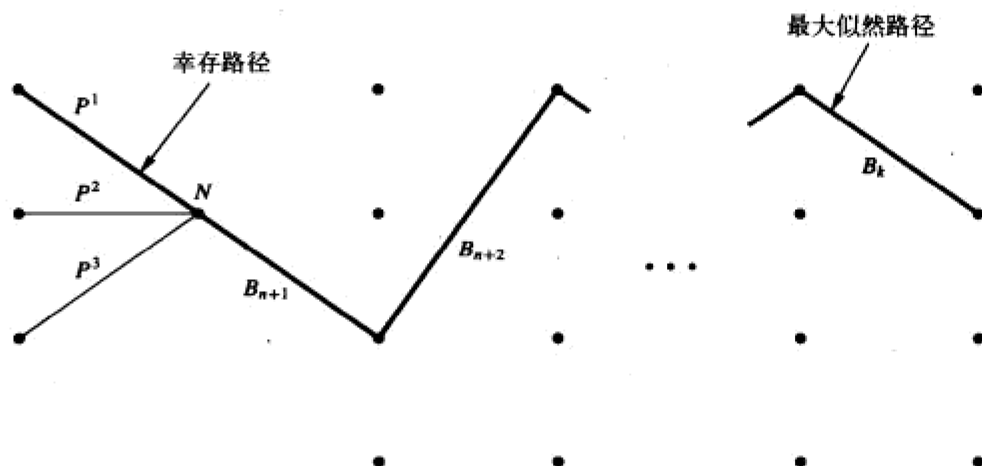


图8-8 最大似然路径上的部分路径度量

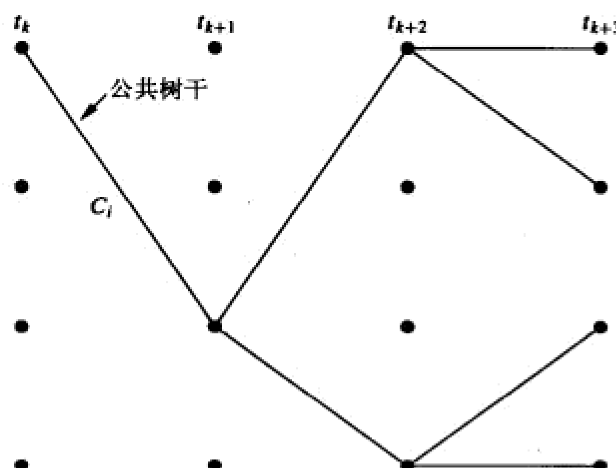


图8-9 所有幸存路径的公共树干

维特比算法巧妙地利用了格图的这种结构特点，在每个节点处，留下到达该节点时局部度量最大的路径，丢弃其余的局部路径。这个留下的路径叫做*幸存路径*（survivor path）。对于图8-8中的例子，路径1是节点N的幸存路径，路径2和路径3都被丢弃。在格图中，每个时刻有2^{K-1}条幸存路径，各对应一个编码器状态。如果所有幸存路径在某个支路处是重合的，就可以输出这个支路上的数据的译码结果。图8-9中示出了t_{k+3}时刻的所有幸存路径，可以看到，所有幸存路径在t_k到t_{k+1}之间有一个共同的树干，因此在t_{k+3}时刻，译码器就可以输出对应t_k到t_{k+1}之间这个分支上的数据C_i。注意，对于某个时刻，退回多少时间能出现重合树干是不固定的，它与k、K码的具体特性都有关系。为了避免随机的译码时延，经常将维特比译码算法做一点修改：在截止到当前的局部路径中，

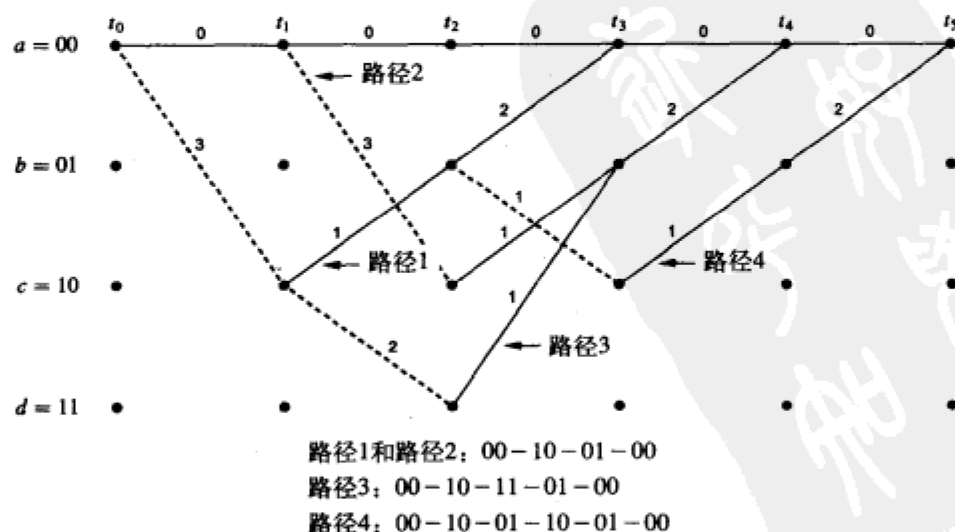
找出局部累积度量最大的，沿着这个路径回退n步会有一个分支，输出这个分支上的数据。这样修改后的算法不再是最大似然译码，但若n足够大的话（一般需要$n \geqslant 5K$），性能也非常接近最大似然译码。

252

维特比算法必须存储$2^{k(K-1)}$条幸存路径及其度量。在每一步，每个节点需要计算2^k个度量，以从2^k个到达路径中选出幸存路径。因此，维特比算法所需要的计算量及存储量随着k和K的增加而呈指数增长。因此实际应用中k和K的值不能太大。

8.3.4 距离特性

与分组码一样，卷积码的纠错能力也与码字序列之间的距离有关。卷积码是线性码，所以编码序列之间的最小距离可通过比较全零序列与其他序列得到，也就是比较格图上的全零路径和其他路径。很明显，与全零路径距离最小的路径必然是先从全零路径离开，后来又回到全零路径。这两个路径有一部分不同，其余都是重合的。为了找到距离最小的路径，必须要考查所有从全零状态出发、之后又回到全零状态的路径。图8-10中画出了图8-7中t_0时刻到t_5时刻之间所有从全零状态出发又回到全状零态的路径。注意到路径2与路径1其实是一样的，只是在时间上进行了移位，所以我们把这两条路径看成是一条。此外，如果我们把观察时间放得更长的话，那么新出现的零始零终路径一定会通过已经画出的某些分支及其延迟，使所画出的路径再增加一些分支，其路径度量只能是更大。比如图8-10中的路径4经历的分支是00－10－01－10－01－00，路径1经历的分支是00－10－01－00，路径4的前两次状态转移和最后一次状态转移与路径1相同，而路径4还有其他的转移，因此路径4的路径度量无论如何也不会比路径1小。这就是说，我们没有必要考虑那么长的时间间隔来寻找最小距离路径。图8-10标出了每条分支和全零序列之间的汉明距离。将每条路径上各分支的汉明距离累加就可以看到，路径1的汉明距离是6，路径3和路径4的汉明距离为8。前面提到，虚线代表的是输入比特为“1”，实线代表的



路径1和路径2：00－10－01－00
路径3：00－10－11－01－00
路径4：00－10－01－10－01－00

图8-10 与全零路径的距离

是输入比特为“0”，因此路径1在t_0到t_5之间的输入比特序列为10000，路径3对应的输入序列为11000，路径4对应的是10100。相对全零序列来说，路径1产生了1个比特错误，路径3或路径4产生了2个比特错误。

253

定义卷积码的最小自由距（minimum free distance）d_f为格图中所有路径与全零路径之间的最小汉明距离，简称自由距。前面这个例子中自由距就是6。卷积码的码字纠错能力和分组码一样，只是把d_{min}换成了d_f，即它能纠正$t=\lfloor 0.5d_f \rfloor$个错误。

8.3.5 状态图和转移函数

卷积码的转移函数能反映所有零始零终的非全零路径，也可用来计算错误率的界。可以从卷积码的状态图得到转移函数。图8-11是图8-7的格图所对应的状态图，我们将全零状态表示成两个节点：节点$a=00$代表起始状态，节点$e=00$代表终了状态。输入比特0引起的状态转移用实线表示，输入比特1引起的状态转移用虚线表示。每个支路标记有$D^0=1, D^1, D^2$或者D^3，其指数表示该支路对应的编码输出与全零码字之间的汉明距离。图中节点a处的自环可以不画，它对卷积码的距离特性没有作用。

254

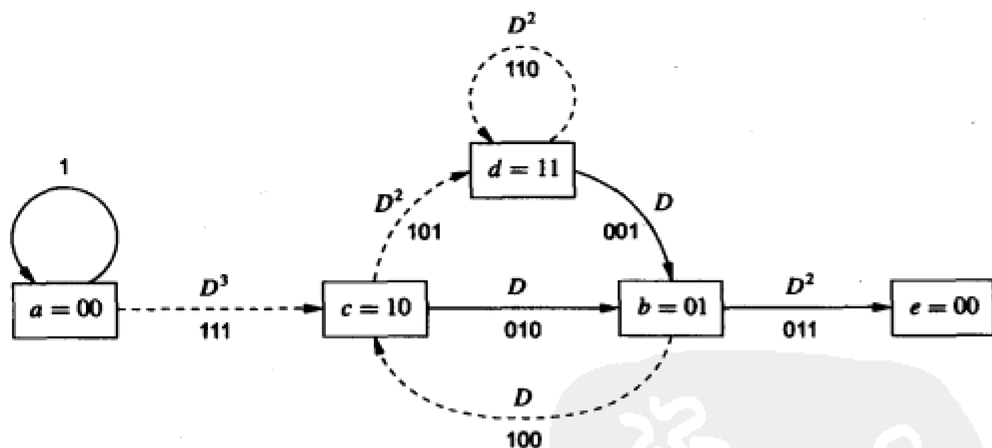


图8-11　状态图

这个状态图可以用每个状态的状态方程表示。对于图8-11，可以写出下面四个状态方程：

$$X_c = D^3X_a + DX_b \quad X_b = DX_c + DX_d \quad X_d = D^2X_c + D^2X_d \quad X_e = D^2X_b \tag{8-61}$$

其中$X_a, \cdots, X_e$是代表局部路径的虚变量。状态a到状态e的路径转移函数定义为$T(D)=X_e/X_a$。采用梅森公式等方法求解状态方程，可得到转移函数有如下形式：

$$T(D) = \sum_{d=d_f}^{\infty} a_d D^d \tag{8-62}$$

其中a_d是距全零路径的汉明距离为d的路径个数。由于距全零路径的最小汉明距离为d_f，所以转移函数$T(D)$表明有a_{d_f}条最小距离的路径。以图8-11为例，求解式（8-61）的状态方程组得到转移函数为：

$$T(D)=\frac{D^6}{1-2D^2}=D^6+2D^8+4D^{10}+\cdots \tag{8-63}$$

这说明有一条最小距离$d_f=6$的路径，有2条汉明距离为8的路径，这个结果与图8-10吻合。转移函数以简洁的方式列举出了不同距离的零始零终的非全零路径个数。

255

尽管转移函数已经足以表征出格图中所有不同距离的零始零终路径的个数，但为了计算卷积码的比特错误概率，还需要更多的细节信息。为此可在转移函数中增加两个参数：输入比特为1时的分支（图8-11的虚线）增加一个因子N，所有分支都增加一个因子J。转移函数中J的指数等于从节点a到节点e的路径所包含的分支数。图8-12是对应于图8-7的扩展状态图。

扩展状态图也可以用状态方程来表达。对于图8-12，我们有

$$\begin{aligned} X_c &= JND^3X_a+JNDX_b \qquad X_b=JDX_c+JDX_d \\ X_d &= JND^2X_c+JND^2X_d \qquad X_e=JD^2X_b \end{aligned} \tag{8-64}$$

扩展状态图的转移函数可类似定义为$T(D,N,J)=X_e/X_a$。对这个例子有：

$$\begin{aligned} T(D,N,J) &= \frac{J^3ND^6}{1-JND^2(1+J)} \\ &= J^3ND^6+J^4N^2D^8+J^5N^2D^8+J^5N^3D^{10}+\cdots \end{aligned} \tag{8-65}$$

考虑有限长度的传输序列时，J是一个重要的参数。对于无限长序列，可令$J=1$而使扩展状态图的转移函数变为：

$$T(D,N)=T(D,N,j=1) \tag{8-66}$$

256

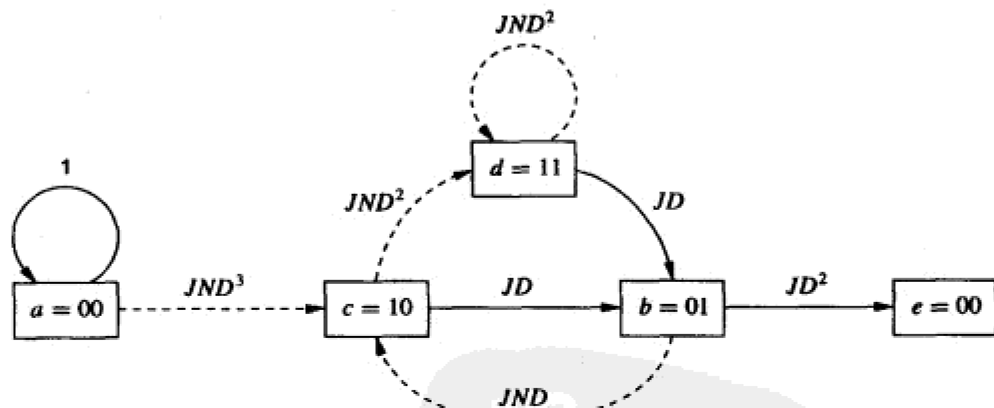


图8-12 扩展的状态图

扩展状态的转移函数包括了更多的信息。例如式（8-65）表明，汉明距离为最小值6的路径长度是3，因为N的指数是1，所以将全零路径误为这个路径将产生1比特的错误；两条汉明距离为8的路径中，有一条长度为4，它将产生2比特错误，另一条长度是5，也将产生2个比特的错误，这些结果都与图8-10的格图吻合。扩展转移函数能方便地表示出零始零终路径的汉明距离、路径长度以及比特错误数。下一节中我们将会看到，传递函数对分析卷积码的错误概率也非常有用。

8.3.6　卷积码的错误概率

卷积码是一种线性码，所以它的错误概率等于发送全零序列时译为其他序列的概率。下面的分析将同时考虑硬判决译码和软判决译码的错误率。由于软判决译码有更好的性能，无线通信系统中广泛使用的是软判决译码。

首先考虑软判决译码。我们关心的是发送全零序列时，译码后得到其他序列的概率。假设编码输出用BPSK调制通过AWGN信道传输，编码符号的能量为$E_c=R_cE_b$。由文献[1]可得，发送全零序列而译码结果是汉明距离为d的一条路径的概率为：

$$P_2(d)=\mathrm{Q}\left(\sqrt{\frac{2E_c}{N_0}d}\right)=\mathrm{Q}\left(\sqrt{2\gamma_b R_c d}\right) \tag{8-67}$$

这个概率叫成对错误率（pairwise error probability），它是一对路径之间的错误概率。转移函数列出了所有零始零终路径，因此全零路径被错译为其他路径的概率的联合界为：

$$P_e\leqslant\sum_{d=d_f}^{\infty}a_d\mathrm{Q}\left(\sqrt{2\gamma_b R_c d}\right) \tag{8-68}$$

式中a_d是距离全零路径为d的路径个数。将切尔诺夫界用于Q函数得

$$\mathrm{Q}\left(\sqrt{2\gamma_b R_c d}\right)\leqslant \mathrm{e}^{-\gamma_b R_c d}$$

利用上式，可将式（8-68）表示为转移函数的形式：

$$P_e<T(D)|_{D=\mathrm{e}^{-\gamma_b R_c}} \tag{8-69}$$

这个上界给出了将一个序列错译为其他序列的概率，但不能反映更为重要的比特错误率。注意到$T(D,N)$中N的指数代表错误路径中信息比特的错误个数，我们将$T(D,N)$表示为：

257

$$T(D,N)=\sum_{d=d_f}^{\infty}a_d D^d N^{f(d)} \tag{8-70}$$

其中$f(d)$是与全零路径距离为d的路径所包含的比特错误数。这样，对于$k=1$的卷积码，比特错误率的上界为（见文献[1]的8.2节）：

$$P_b\leqslant\sum_{d=d_f}^{\infty}a_d f(d)\mathrm{Q}\left(\sqrt{2\gamma_b R_c d}\right) \tag{8-71}$$

它与式（8-68）的差别只是加权因子$f(d)$。采用Q函数的切尔诺夫界可得：

$$P_b<\left.\frac{\mathrm{d}T(D,N)}{\mathrm{d}N}\right|_{N=1,D=\mathrm{e}^{-\gamma_b R_c}} \tag{8-72}$$

将式（8-71）或者式（8-72）除以k可得到$k>1$时的误比特率。

上面的分析是针对BPSK调制的，它也适用于QPSK调制（QPSK等效于两个独立的BPSK）。对于其他调制方式，需根据该调制的误码率公式重新计算成对错误率$P_2(d)$。

再来考虑硬判决译码。发送全零路径时，若距离d处的路径离接收序列更近，就会

错译为这个路径。若d为奇数，则

$$P_2(d)=\sum_{k=0.5(d+1)}^{d}\binom{d}{k}p^k(1-p)^{(d-k)} \tag{8-73}$$

其中p是信道的错误率。上式是因为，至少要发生0.5（$d+1$）个错误才会导致译码器选择错误的路径。当d为偶数时，若错误比特数超过$0.5d$个，译码器将译错，若错误比特数恰好等于$0.5d$，译码器将随机选择。

成对错误率的上界为：

$$P_2(d)<[4p(1-p)]^{d/2} \tag{8-74}$$

类似于软判决译码时的情形，硬判决译码的错误率的界为：

$$P_e<\sum_{d=d_f}^{\infty}a_d[4p(1-p)]^{d/2}<T(D)|_{D=\sqrt{4p(1-p)}} \tag{8-75}$$

$$P_b<\sum_{d=d_f}^{\infty}a_d f(d)P_2(d)=\left.\frac{dT(D,N)}{dN}\right|_{N=1,D=\sqrt{4p(1-p)}} \tag{8-76}$$

258

8.4 级联码

级联码如图8-13所示，它是由内码和外码组成的两级编码。一般靠内码纠正大部分的信道差错，剩余的错误靠纠错能力稍弱一些的外码克服。级联码能有效对抗无线衰落信道中常常出现的突发错误。在低信噪比时，卷积码经过维特比译码后经常出现突发错误。由于RS码纠突发错误的能力强，所以为了补偿这些突发错误经常的用法是以卷积码为内码，再级联一个RS码作为外码。级联码设计常常在外码和内码之间设一个交织器，以打散突发错误。我们将在8.8节中介绍交织器的设计问题。

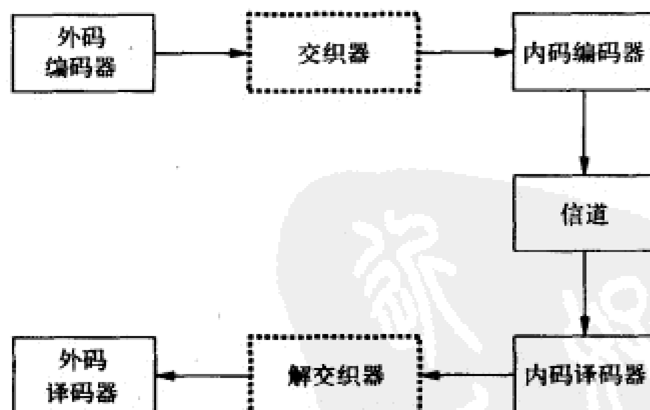


图8-13 级联码

相同错误概率时，级联码的复杂度一般比单一的码更低。级联码的译码分为两步，先译内码，再译外码。这是次佳的方法，最佳方法应该是内码和外码一起进行最大似然

译码，不过这样做的复杂度太高。20世纪90年代中期出现了一种迭代译码方法，它能使级联码的译码接近最佳，这就是下面要讲的Turbo码。

8.5 Turbo码

1993年，Berrou、Glavieux和Thitimajshima在其经典文献[6~7]中提出了一种性能和香农极限只差零点几个分贝的高性能编码，即Turbo码。此后，Turbo码及更为一般化的基于图的具有迭代译码算法的编码[8-9]受到了人们的广泛关注，这些码的某些特性至今尚未被人们完全理解。早在20世纪60年代初，Gallager就已经提出了基于图的编码思想[10]。当时人们认为这样的编码很难实现，因此没有引起多数研究者的重视。1993年那篇Turbo码的经典文献[6]激发了人们的热情，开始重新考察Gallager的工作以及其基于图的迭代译码技术。

259

Berrou等人首先提出的这种Turbo码包括两个关键的部分：一是并行级联编码，二是迭代译码，也即“Turbo”译码[6, 11]。图8-14是典型的并行级联编码器，它包括两个由一个交织器分开的并行的卷积编码器。编码输出包括数据比特m及校验比特X_1、X_2。数据直接出现在编码结果中，所以它是一个系统Turbo码。并行编码中起关键作用的是递归卷积码和交织器。Turbo码一般存在背景误码（error floor）[1]现象，交织器的设计对此有重要影响[11]。

260

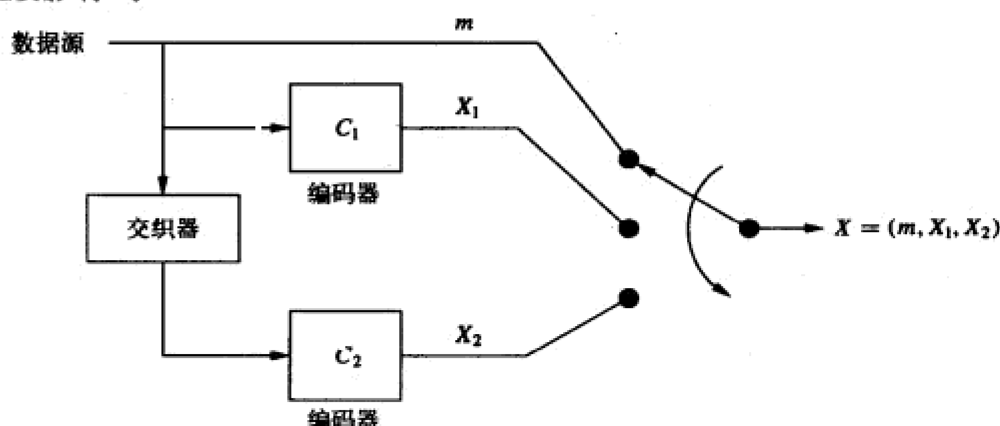


图8-14 并行级联码（Turbo码）的编码器

与编码器结构相对应，译码器也由两个成员译码器组成。译码器利用这种结构进行迭代译码，也称作Turbo译码。图8-15是对应于图8-14的译码器，其中每个成员译码器都是软入软出的。译码器1根据接收到的（m, X_1），输出一个关于发送比特概率$p(m_1)$的软判决值。这个关于概率$p(m_1)$的测度值可通过MAP（maximum a posteriori）算法或者SOVA（soft output Viterbi algorithm）算法产生。SOVA是对维特比算法的扩展，它给普通的维特比算法附加了一个可靠性度量（见文献[4]的12.5节）。译码器1将这个概

1. 许多关于Turbo码的中文文献中有时也称其为误码率的地板效应或者误码平台。本文中的译法参考了noise floor，一般将其译为背景噪声或本底噪声。——译者注

率信息送到译码器2，译码器2根据接收到的（m, X_2）及$p(m_1)$产生出它自己的概率测度$p(m_2)$。然后把$p(m_2)$送到译码器1，译码器再结合原始的接收码字修正概率测度$p(m_1)$。修正后的$p(m_1)$又送到译码器2，译码器2再修正$p(m_2)$。如此往复，两个译码器交替更新它们的输出概率值。理想情况下两个译码器最终将给出一致的概率输出，从而形成硬判决$m=m_1=m_2$。不过在实际中，迭代译码应该在何时中止还是一个问题，部分原因是因为Turbo译码算法在许多情况下并不收敛，也就是说两个译码器对信息m的判断并不一致。监测是否收敛的方法有好几种，包括方差估计[7]及一些基于神经网的技术[12]。

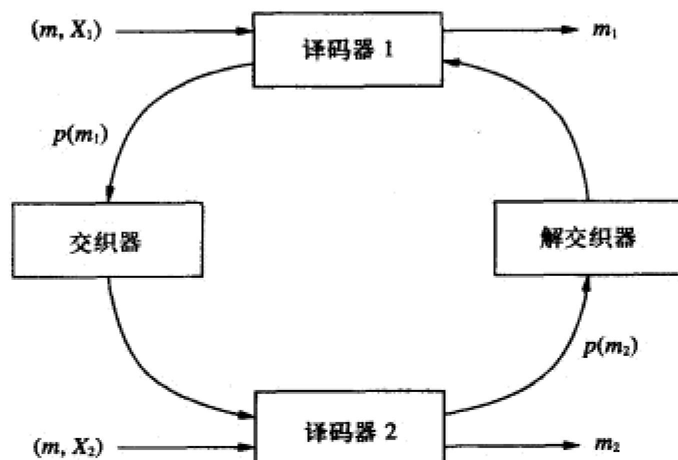


图8-15 Turbo码的译码器

图8-16示出了不同迭代次数下Turbo码的仿真性能。该Turbo码是由两个编码率为1/2、约束长度为$K=5$的卷积码构成，交织器长度是$d=2^{16}=65536$。从图中可以看出，

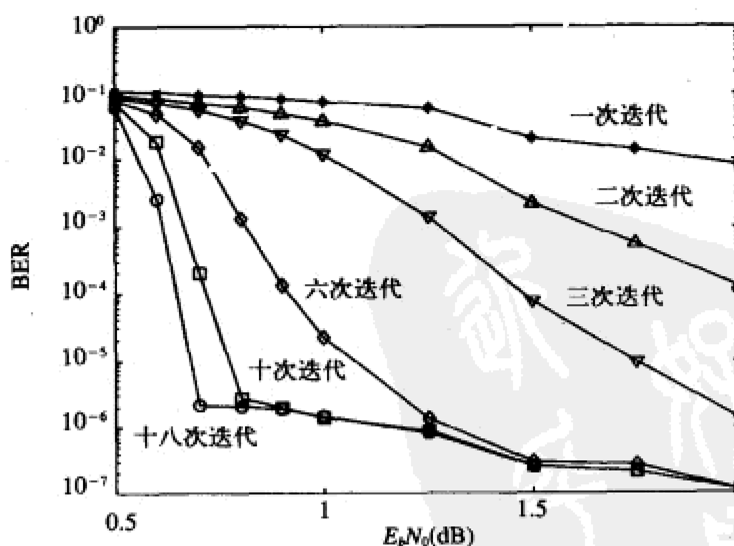


图8-16 Turbo码的性能（码率1/2、成员码$K=5$、交织长度2^{16}）

译码器在大约18次迭代后基本收敛。图中体现了Turbo码的几个重要特性。首先可以注意到Turbo码突出的性能：达到误比特率10^{-6}所需要的E_b/N_0小于1dB。实际上，文献[6]中最先提出的那个Turbo码在误比特率为10^{-5}时，对应的E_b/N_0离香农极限只差0.5dB。这个神奇的性能是因为Turbo码的编码复杂度已经足够达到香农极限，而迭代译码使译码过程不至于过分复杂。从这个图中还也可以看到Turbo码的背景误码现象，发生在10^{-6}处。如果系统要求极低的误比特率，这将是一个很大的问题。存在一些能够降背景误码的方法，包括改进交织设计、增加成员码的约束长度等。

261

并行级联也可以改成串行级联[13]。串行级联编码中，第一个成员编码器是外码，它的输出经交织后送往作为内码的第二个编码器，其输出送往信道。内码和外码之间采用迭代译码。有许多文献研究了串行级联码和并行级联码的性能[13-15]。总的来说，在时延及复杂度相当的情况下，这两种码的性能都很好。某些情况下，串行级联码在很低误比特率时性能更好，背景误码也比较低。

8.6 低密度校验码

低密度校验码（low-density parity-check，LDPC）码和Turbo码在有些地方有相同的思想。它最早是由Gallager发明[10]的，但一直被人们遗忘。1993年那篇Turbo码的标志性论文[6]发表之后，Mackay、Neal[16]以及Wiberg[17]提出了LDPC码。人们很快发现，这实际上是Gallager的原始思想的重新提出。此后的许多工作致力于容量极限、编码和译码设计、不同信道下的实际实现等问题。

LDPC是一种有特殊校验矩阵$\boldsymbol{H}$（见8.2.3节）的线性分组码。对于参数为(d_v, d_c)的规则二进制LDPC码，校验阵$\boldsymbol{H}$的每一列都有d_v个1，每一行都有d_c个1。这两个参数相对于码长来说非常小，因此$\boldsymbol{H}$阵中非零的元素很少，即1的密度很低，这是它称为低密度校验码的原因。

LDPC码在码字足够长时也能逼近香农极限，有时性能比并行或串行级联码更好[18]。从实现的角度看，Turbo码和LDPC码的区别在于：Turbo码的编码复杂度低，但随码长线性增长，译码复杂度提高很多，译码复杂度主要是因为迭代和消息传递。与此相反的是，LDPC码的编码复杂度高，译码复杂度低。和Turbo码一样，LDPC的译码也采用了人工智能领域中常用的Pearl置信度传播算法[19]。但LDPC中的置信度传播比Turbo码中的简单[19-20]，并且置信度传播算法可以并行实现，也能用低复杂度的译码器很好的近似[21]，当然Turbo码也可以这样做。另外，LDPC码能检测出译出的码字是否正确，Turbo码则不一定能做到这一点。如何在LDPC和Turbo码之间权衡是当前一个很活跃的话题，还存在许多待研究的问题。

LDPC码方面的研究工作还包括容量极限[21]、高效的编码设计[22]、高效的编码和译码算法[21-23]、非规则码[18]、多进制码[24]、编码调制[25]等。

262

8.7 编码调制

虽然香农早在20世纪40年代末就证明了AWGN信道的容量理论，但是直到20世纪

90年代才在二进制调制的AWGN信道下真正获得了逼近香农极限的速率。香农的理论预言，通过编码我们能够同时减少能量和带宽，但如同8.1节所指出的，传统的纠错码方案，如分组码、卷积码、Turbo码等，无一不是以增加带宽或者减少数据率为代价来获得编码增益的。

高频谱效率编码方面的突破是Ungerboeck所提出的编码调制技术[26]。它将信道编码和调制进行联合优化，从而无需花费额外的带宽，就能获得显著的编码增益。Ungerboeck最早提出的格码调制通过子集划分将卷积码输出映射到多电平调制或多相调制上，它并不差于后来提出的许多编码调制方案，如陪集码、点格码以及更复杂的网格码[27]。下面我们将简要介绍编码调制的一般原理。更多的内容可参考文献[26~28]。

图8-17示出了基于网格码和点格码，或者更一般地说是任何形式的陪集码的编码调制的实现框图，包括以下五个基本步骤：

（1）一个二进制的卷积码或分组码编码器E，它将k个数据比特编成$k+r$个编码比特；

（2）一个子集（陪集）选择器，它用编码比特从N维星座的2^{k+r}个子集中选出一个子集；

（3）一个星座点选择器，它用$n-k$个未编码比特从所选子集中的2^{n-k}个星座点中选出一个星座点；

（4）一个星座映射器，它将所选的N维星座点映射为连续$N/2$个二维星座点；

（5）一个MQAM调制器或者其他的M进制调制器。

263

其中的前两步是编码，后三步是调制。接收端基本上是以相反的次序操作。首先是MQAM解调，然后是二维到N维的星座映射。译码过程基本分为两部分：首先将接收信号点判决为每个子集中距离最近的星座点，然后计算最大似然的子集序列。如果E是卷积编码器，这样的编码调制方案称为*网格码调制*（trellis coded modulation），如果E是分组码编码器，则称为*点格码调制*（lattice coded modulation），也简称为格码或点阵码。

上述五个步骤基本上可以分出信道编码带来的增益和调制星座成形带来的增益。子集划分基本上可以从星座成形中分出，这样子集划分和编码器（E）的特性决定了信道编码的距离特性，进而决定了信道编码的增益。下文中我们将详细讨论信道编码增益。当N趋于无限大时，最优的信号星座成形能带来1.53dB的成形增益，这个增益与具体采用的信道编码无关[1]。成形增益付出的代价是星座映射的复杂度随N指数增长。星座的大小由平均功率约束条件决定，不影响成形增益或编码增益。

可以将N个比特组成的序列看成是N维空间或称*序列空间*（sequence space）中的一个点，这样，信道编码就是在N维空间的所有可能点中选择出了一些点，如果选择适当，使得不同输入序列之间有一定的最小距离d_{min}的话，就能产生出编码增益。经常发生的错误是将发送序列误译为离它最近的序列，这样的错误概率在AWGN信道中随d_{min}^2的增加而降低。因此，若要降低误码率，就需要增加序列空间中点与点之间的距离。直接拉

1. 正方体星座的成形增益为0dB，球体在相同体积下有最小的能量，所以当N给定时，球形星座能获得最大的成形增益，详见文献[27]。

伸序列空间能使空间中各点之间的距离平方统一扩大一个固定的比例，从而能降低错误率，但这也使发送功率增加相同的比例，因此直接拉伸不会带来编码增益。因此，有效的功率增益应当是编码产生出的所有序列点之间的最小平方距离乘上所选序列点的密度。假设序列空间本身的最小距离和密度分别为d_0和Δ_0，通过编码选择的序列空间，其最小距离和密度分别为$d_{\min}$和Δ，那么在最大似然序列检测时可获得的信道编码增益为：

$$G_c = \left(\frac{d_{\min}^2}{d_0^2}\right)\left(\frac{\Delta}{\Delta_0}\right) \tag{8-77}$$

式中的(Δ/Δ_0)也称为*星座扩展因子*（constellation expansion factor）。若编码器E的冗余比特数为r，则每N维的扩展因子是2^{-r}[27]。

264

在选择出的序列空间中，如果序列存在多个最近邻的话，译码出错的可能性就会大一些，这个因素将使式（8-77）给出的标称编码增益打一些折扣，损失的大小用*错误系数*（error coefficient）描述。常见点阵码调制及格码调制的错误系数见文献[27]。通常在G_c相当时，点阵码的错误系数一般比格码大。

信道编码是通过对点阵进行子集划分实现的。*点格*（lattice）是欧几里德N-维空间中的一些矢量构成的离散集合，它对于普通的矢量加法运算是一个群。因此点格中任意两个矢量的和及差还在这个点格中。子格是点格的一个子集，它本身也是一个点格。由未编码的MQAM调制的星座点序列构成的序列空间是N维立方体[1]，其最小距离和二维时一样。可以让编码输出序列落在序列空间中的一个点格内，如果这个点格比N维立方体更密，那么我们就可以增加$d_{\min}$而保持密度不变，这样做并没有扩大星座。对于$N>1$，存在比N维立方体更密的点格。在N维中寻找最密的点格是一种数学问题，其解对任意N存在，相应的译码复杂度也可以接受[2]。找到最密的点格后，可通过对任意的子格进行变换而得到陪集，这样形成了对点格的子集划分。子格的选择决定了划分的大小，也就是图8-17中子集选择器的可选子集数目。这样，数据比特是以两种方式传送的：陪集序列携带了一部分比特，每个陪集中的星座点携带了其余的比特。点格的密度决定每个陪集中星座点之间的距离，二进制编码器E的冗余比特数r及其他编码特性决定了陪集序列之间的距离。若记d_p为陪集中星座点之间的最小距离，d_s为陪集序列之间的最小距离，那么最小码距是$d_{\min}=\min(d_p, d_s)$，有效编码增益为：

$$G_c = 2^{-2r/N} d_{\min}^2 \tag{8-78}$$

其中的$2^{-2r/N}$是二维空间中的星座扩展因子。

再回到图8-17，假设我们需要发送mN个比特，每维$m=n+r$比特。采用最密的点格，它落在一个N维球体内，球体半径范围内覆盖了2^{mN}个点。点格的密度和编码器特性产生了编码增益，N维球体相对于N维立方体产生了成形增益，这两个单独的增益合起来就是总增益。增大球的半径（即增加信号功率）可以使球容纳更多的点，它直接增加了陪集中的星座点数，因此就能传送更多的未编码比特。$N=4$、8、16和24时，对

1. 坐标点为奇整数的二维矩形点格的笛卡儿积。
2. 采用维特比算法时，最大似然译码的复杂度大致与N成正比。

二维矩形格的简单划分，再结合传统的分组或者卷积码，就可以产生出最大密度的N维点格。这种构造方式及相应的译码算法见文献[28]，它同时适用于点格码及网格码。点格码在$N=4$、8、16和24时，相应可获得的有效增益大约为1.5dB，3.0dB，4.5dB和6dB。复杂度相当的网格码能获得更大的增益。

265

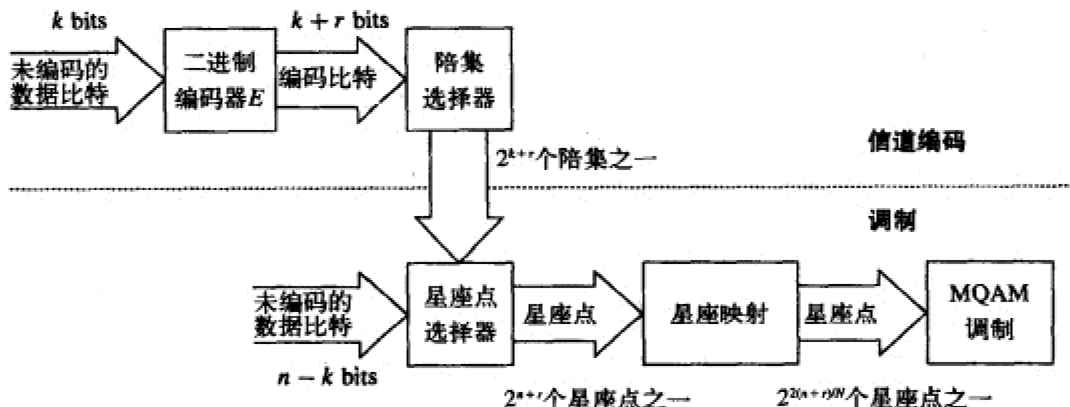


图8-17 编码调制的一般形式

我们来看文献[28]中给出的一个点格码的例子，其$N=8$、增益为3dB。首先按图8-18所示的方式将二维星座划分成四个子集，分别用A_0、A_1、B_0和B_1表示。比如16QAM星座的四个子集中每个包括四个星座点。每个子集内星座点之间的距离是原来的两倍。8维点格由4个这样的星座点构成的序列组成，每个序列中的星座点要么全是A点，要么全是B点，且A点或B点的下标满足校验方程$i_1+i_2+i_3+i_4=0$，就是说序列的下标是一个（4，3）偶校验码的码字，其最小汉明距为2。三个数据比特和一个奇偶校验比特决定点格的子集，其余数据比特决定子集中的星座点。经过这种子集划分后的最小距离的平方是未编码星座的四倍，即有6dB的增益。不过，校验比特使星座图每维扩展了1/2个比特，根据5.3.3节，相应的功率因子为$4^{0.5}=2$，即3dB。于是，净编码增益是$6-3=3$dB。如果数据率为每符号m比特，那么点格的四个子集中每个都有2^{m-1}个点[1]。例如16QAM中，四个子集每个都有$2^{m-1}=4$个点，数据率是每个16QAM符号含3个数据比特。

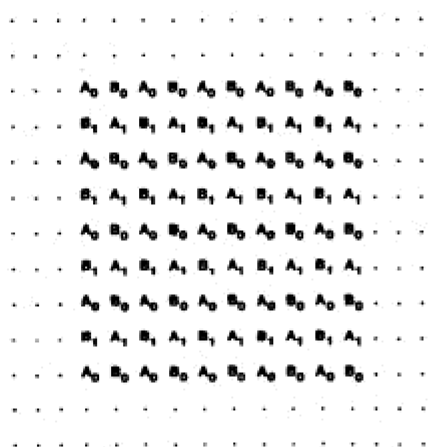


图8-18 8维点格的子集划分

266

文献[29~31]研究了采用Turbo码的编码调制，结果表明Turbo网格码调制能接近多进制调制的香农极限。

1. 即每个符号携带$m-1$个比特，另一个比特体现在编码中。

8.8 衰落信道下的编码和交织

前述的分组码、卷积码及编码调制都是为AWGN信道设计的。在衰落信道中，深衰落会使解调输出发生突发错误，而多数为AWGN信道所设计的码都不能纠正长突发错误，这使它们在衰落信道中表现不佳。

为了克服突发错误的影响，衰落信道中一般是把编码与交织结合起来运用。其原理是通过交织器把长突发错误分散到各个码字中，每个接收码字中的错误个数很少，在码字的纠错能力之内。这样突发错误能通过交织器来打散，而纠错可以通过码字来完成。交织器的长度必须足够大，以使接收码字中的衰落呈现为独立。慢衰落信道需要更长的交织器，相应会有更大的时延。

交织编码是一种分集，其性能可用误码率的分集阶数来衡量，它一般是最小汉明距的函数，因此，衰落信道中交织编码的设计应当使分集阶数最大化。而在AWGN信道中，设计的主要考虑是让欧氏距离最大化。下面的小节将详细讨论用于分组码、卷积码和编码调制的交织编码。我们假设接收端已知信道衰落的信息，这个假设将大大简化分析过程和译码器设计。一般可通过发送导频符号来获得信道的衰落信息[32-33]。无信道衰落信息时的最大似然译码计算量非常大[34]，故此一般需要对最大似然译码度量或者信道进行近似处理[34-39]。8.5节讲到的Turbo码内部已经有一个交织器，不过这里的交织器设计与衰落信道下的交织器设计有不同的考虑。Turbo码在相关衰落信道下的性能分析及交织器设计可参考文献[40]的第8章及文献[41，42]。LDPC码也可以经过特别设计以具有对抗衰落的分集能力，其性能和针对衰落信道优化过的Turbo码类似[43]。

267

8.8.1 分组码的交织

分组码一般采用分组交织来打散衰落带来的突发错误。分组交织器是一个d行n列的阵列，如图8-19所示。(n, k)分组码的码字按行读入交织器，每行为一个码字，再按列读出送到调制器，然后再通过信道发送。同一码字中的编码符号在传输过程中相隔了$d-1$个别的符号，如果这个时间间隔大于信道的相干时间，即若$dT_s>T_c\approx 1/B_D$，那么同一码字中的码字符号所经历的衰落可以近似认为是独立的，这里的T_s是码字符号的持续时间、T_c是信道的相干时间、B_D是信道的多普勒频移。满足$dT_s>T_c$的交织器称为深交织器。解交织器是一个与交织器相同的阵列，从解调器来的比特按列读入，每行为一个可能包括信道差错的码字，再按行逐码字读出到译码器。

从图8-19中我们可以看到交织编码纠正突发错误能力。假设二进制(n, k)分组码的纠错能力为$t=2$，信道中突发错误的长度为3。当码字直接传输时，接收的4个码字中有3个错误，超过了纠错能力，将被译错。当码字通过图8-19所示的交织器/解交织器传输时，这个突发错将被分散到三个码字中，每个码字有1比特错误，三个码字都能正确译码。8.8.2节将要提到的卷积交织器在概念上与此相同，它更适合于卷积码。

268

分析交织编码的性能需要计算成对错误率或者利用切尔诺夫界或联合界，具体分析方法见文献[1]的14.6节。联合界给出了性能的简单估计。考虑采用深交织的瑞利衰落信

道，各编码符号的衰落近似独立，则对于非相干FSK调制的（n, k）分组码及软判决译码，码字错误率的联合界为

$$P_e < (2^k - 1)[4p(1-p)]^{d_{\min}} \tag{8-79}$$

其中$d_{\min}$是最小码距，p为

$$p = \frac{1}{2 + R_c \bar{\gamma}_b} \tag{8-80}$$

对于慢变衰落信道，接收端可以获得相干参考相位，故可采用BPSK调制，此时（n, k）分组码在软判决译码下的码字错误率的联合界为

$$P_e < 2^k \binom{2d_{\min} - 1}{d_{\min}} \left(\frac{1}{4R_c \bar{\gamma}_b}\right)^{d_{\min}} \tag{8-81}$$

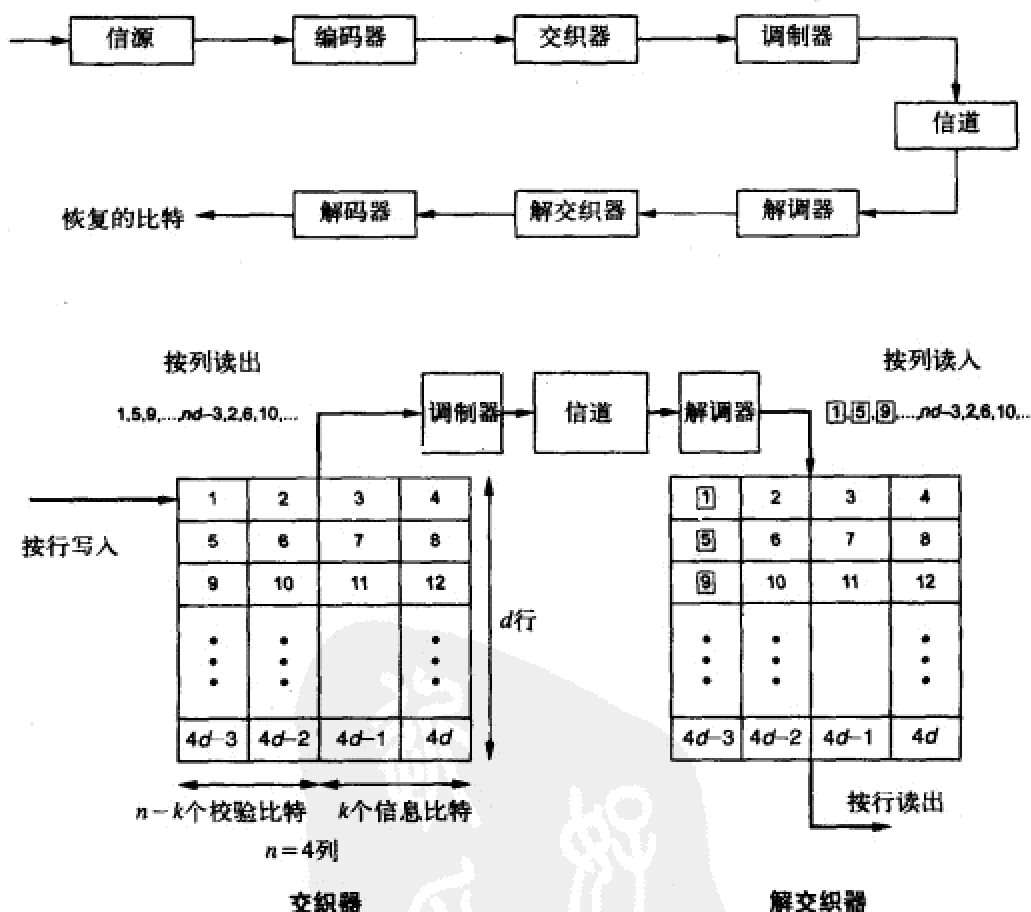


图8-19 交织/解交织

注意式（8-79）和式（8-81）的形式与MRC合并的误码率公式（7-23）相似，$d_{\min}$可以看成是分集阶数。硬判决译码时也可以得到类似的结果，只不过分集阶数要小一半。由

此可见，为了优化性能，衰落信道中分组交织编码的设计应该使汉明距最大化。

对于真正的最大似然译码来说，衰落信道中的相关性包含了有用的信息。交织编码丢弃了这个信息，潜在降低了信道的固有容量文献[44]，所以交织编码并不是最优的。抛开这一点不谈，由于符号相关时的最大似然译码复杂度不可承受，所以常用的方式还是对AWGN信道中编码进行交织。

269

例8.8：考虑一个多普勒频移$B_D=80\text{Hz}$的瑞利衰落信道。系统采用（5, 2）汉明码和交织器来对抗衰落。如果码字符号的传输速率是30kbit/s，试求能使符号衰落独立所需要的交织深度。此时可对抗的最大突发错误长度是多少？并求出对应的交织时延。

解：（5, 2）汉明码的最小码距为3，它能纠正$t=[0.5\times3]=1$个码字符号错误。码字符号的传输速率为30kbit/s，符号周期为$T_s=1/R_s=3.3\times10^{-5}$。信道的相干时间为$T_c=1/B_D=0.0125\text{s}$。交织后符号之间的间隔时间是$dT_s$，欲使每个码字符号所经历的衰落为独立，需$dT_s>T_c$。由此得到$d\geqslant T_c/T_s=375$。突发错误经过交织后的间隔是$d$，如果突发长度为$d$，解交织后每个码字只有一个符号错误，可以被纠正。因此本例可对抗的最大突发错误长度为375符号。交织器按列读出之前，所有的行必须要填满，所以交织时延是$ndT_s=5\times375\times3.3\times10^{-5}=62.5\text{ms}$。这个延时会影响话音系统的质量。可见，交织编码为了能纠正长突发错误，付出的代价是较大的时延。

8.8.2 卷积码的交织

和分组码一样，卷积码的设计并没有考虑纠正突发错误，因此它在衰落信道下的性能会变差。为此，一般也采用交织器来打散突发错误。分组码中，交织器是将突发错误分散到不同的码字中。卷积码没有码字的概念，所以采用了一种略有区别的交织器。这种交织器一般都是用在卷积码中，所以叫卷积交织器。它一方面能打散突发错误，同时又能和编码过程协调一致[45-46]。

卷积交织器的框图见图8-20。编码器的输出分为多路送往多个缓冲器，这些缓冲器的缓冲长度从0开始递增到$N-1$。多路缓冲器的输出复用为一路后通过信道传输。接收端则进行相反的操作。卷积交织器将使编码输出有不同的时延，图中从上到下，交织器各路的时延递增。解交织的时延则与此相反。编码器的连续输出经过交织器送往信道时，出现了长度为$N-1$符号的时间间隔，从而分散了信道中的突发错误。注意卷积交织器也可以用于分组码，但多用于卷积码中。卷积交织器需要的存储量是$0.5N(N-1)$，交织时延是$N(N-1)T_s$（见文献[2]的8.2.2节），T_s是信道传输的符号间隔。

卷积编码和卷积交织下的错误率分析见文献[1]的14.6节。在与先前分析分组码时相类似的条件下，通过切尔诺夫界得到的结论是，软判决译码的分集度数为码的最小自由距，硬判决译码的分集阶数是其一半。

270

例8.9：假设信道的相干时间为$T_c=12.5\text{ms}$，编码比特的传输速率是$R_s=100\text{kbit/s}$。如果编码器输出的相邻符号经过交织后衰落独立，求卷积交织器的平均时延。

解：卷积交织器将已编码比特用NT_s分隔开。为了衰落独立，需要$NT_s\geqslant T_c$，其中$T_s=$

$1/R_s$。因此$N \geqslant T_c/T_s = 0.0125/0.00001 = 1250$。注意，为了达到独立衰落，卷积交织和分组交织对交织深度的要求是一样的。本例中的总时延为$N(N-1)T_s = 15$s，这个时延不论是对语音还是数据来说都是非常大的。

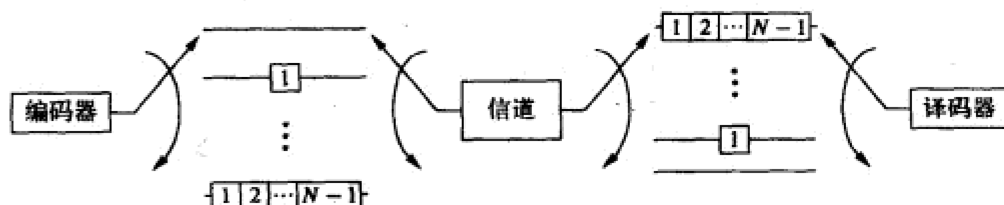


图8-20 卷积交织

8.8.3 符号/比特交织的编码调制

如同分组码、卷积码一样，为AWGN信道所设计的编码调制直接用在衰落信道中时的性能也很差，因此产生了使用交织的想法。不同的是，编码调制下的交织可以有两种选择，第一种方法是将比特进行交织，再映射为调制符号，这种方法称为*比特交织编码调制*（bit-interleaved coded modulation，BICM）。第二种方法是先进行联合的调制和编码，再经过交织后传输，这种方法称为符号交织编码调制（symbol-interleaved coded modulation，SICM）。

初看起来，SICM似乎应该更好一些，因为它的编码和调制是联合进行的，而这正是编码调制设计中核心的理念。不过衰落信道下的编码调制的设计原则需要有所改变，因为衰落信道下的性能取决于码的分集特性，它取决于汉明距，而不是欧氏距。早期对衰落信道中编码调制的研究主要集中在如何提高SICM的分集上，但是很难找到好的设计准则，所得到的性能也不尽人意[47-49]。

BICM[50-51]的提出是衰落信道下编码调制设计中的一个重要突破。BICM的分集度等于错误事件中最小的不相同比特数，而不是信道符号数。它在符号映射前对编码器输出进行比特级的交织，接收端产生出适当的软判决比特度量，再由维特比译码器进行译码。尽管BICM没有遵从调制与编码联合设计的模式，但它的性能却比SICM好很多。而且，BICM的性能分析方法及设计原则也是清楚的[50]。BICM已经成为衰落信道下编码调制的主流技术。

271

8.9 不等差错保护编码

当所传输的不同比特有不同的优先级，或者对误比特率有不同的要求时，可以采用多等级编码或者叫*不等差错保护*（unequal error protection，UEP）编码。例如在同时传送语音和数据的系统中，语音和数据对误码的容忍程度不同：数据包发生错误的话就必须要进行重传，故此数据通常要求$P_b < 10^{-6}$，而高质量的语音所要求的误码率也只在$P_b < 10^{-3}$数量级上。这种情形也出现在信源压缩中，例如在图像压缩中，低分辨率

的比特对重现图像是必须的，而高分辨率的比特只是让图像更清晰。采用多等级的信道编码时，信道好的时候，所有比特都能以很高的概率被正确接收，但在信道处于深衰落的时候，只有高优先级的比特或者要求低P_b的比才能以很高的概率被正确接收。

Imai和Hirakawa[52]最早研究了UEP码的实现问题。随后的文献中开始考虑将二进制UEP码应用到语音信道联合编码[53]和图像信道联合编码[54]中。这些方法采用传统的纠错码（分组码或者卷积码），编码增益直接对应于带宽扩展。此后又提出了两种高频谱效率的UEP码：时分复用的高频谱效率的编码调制[55]以及采用均匀及非均匀星座的编码调制[56-57]，这些多电平编码的设计可以专门针对AWGN信道，也可以专门针对衰落信道。下面我们将对UEP进行简要介绍。主要介绍多电平编码、多级译码的基本原理，包括更为复杂的高频谱效率实现方式。

多电平编码器的结构如图8-21所示。将信源编码器的输出按其优先级大小分为M个并行的比特流，对应设置M个不同码距的二进制纠错码$C_1,\cdots,C_M$。优先级为i的比特由第i个编码器进行编码，输出为s_i。将2^M个点依次从0编号到2^M-1，星座点选择器选出的星座点是

$$s=\sum_{i=1}^{M}s_i\cdot 2^{i-1} \tag{8-82}$$

272

例如，对于$M=3$、8PSK星座，选出的星座点的相位将是$2\pi s/8$。

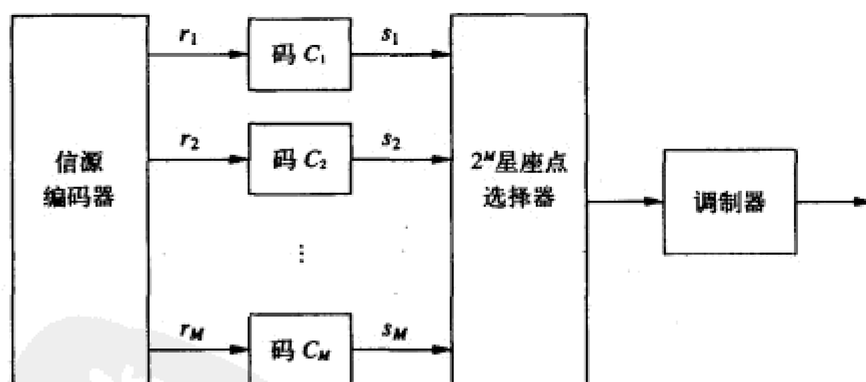


图8-21　多电平编码

多电平编码的最大似然译码是找出能使接收序列概率最大的发送序列，它需对序列$\{s_1\},\cdots,\{s_m\}$进行联合译码。即便每一路编码都很简单，联合译码的复杂度也明显要大大增加。比如说每一路都是状态数为2^{μ_i}的卷积码，$i=1,\cdots,M$，那么联合译码的状态数就是$2^{\mu_1+\cdots+\mu_M}$。因为最优译码的复杂度太大，所以实际中多采用文献[52]提出一种次优方法，叫多级译码。多级译码对多路数据依次进行译码，先译纠错能力最强的C_1，再译纠错能力次强的C_2，依此类推，直到C_M。每次译出C_i后，以后的译码就认为先前的译码结果是正确的。

多电平编码中的二进制编码需要有较多的冗余比特才能产生编码增益，因此频谱效

率不高。文献[56]中提出了一种对8.7节中的编码调制进行时分复用的方案。它对不同优先级的输入数据采用不同编码增益的不同网格码调制或点格码调制。不同数据流的星座大小也有可能不同，但每个星座的平均功率相同。把各路输出的星座点时分复用为一路，再通过信道传输，如图8-22所示。这个图中有两路数据流，对应两种优先级。令R_i（$i=1, 2$）表示图中编码器C_i的比特率，T_1、T_2分别表示高优先级的C_1和低优先级的C_2占用的传输时间，那么总比特率就是$(R_1T_1+R_2T_2)/(T_1+T_2)$，其中高优先级比特所占的百分比是$R_1T_1/(T_1+T_2)$。

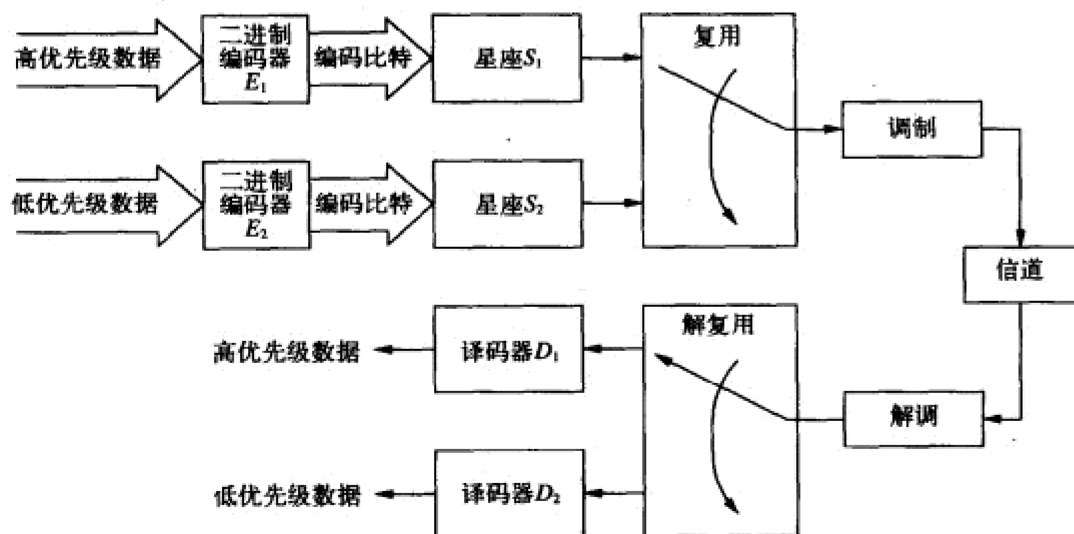


图8-22 时分复用的编码调制

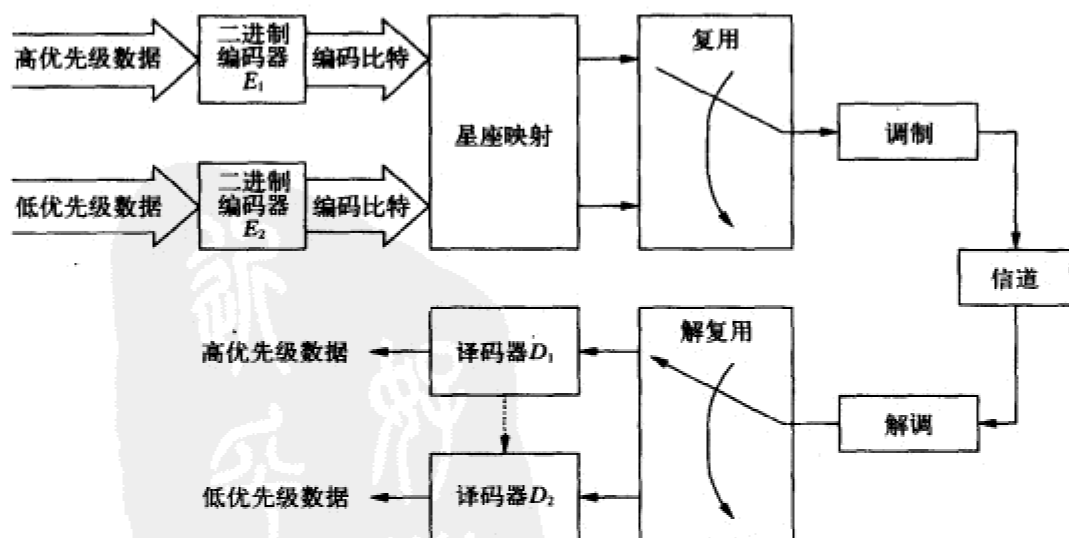


图8-23 信号星座的联合最优化

将图8-22中的星座图S_1和S_2进行联合设计可以获得更大的增益。图8-23就是相应的

改进方案，它可以直接扩展到M路的情形。在网格码调制或点格码调制中，编码比特用来选择子集，未被编码比特用来选择子集内的星座点。二进制的编码器特性只能降低编码比特的错误率P_b，未编码比特的错误率P_b由星座点的距离决定。因此我们可以这样设计来得到两种级别的编码增益：对高优先级的比特采用较强的编码，用它来选择星座划分的子集，对低优先级的比特不编码或者只做较弱的编码，用它来选择星座点。

273

8.10 信源信道联合编码

UEP码的基本思想就是信道编码的误比特率应该和发送比特的优先级或误比特率要求相匹配。发送比特一般是对原始数据进行某种压缩后的结果。所以，UEP编码可以看作是一种信源信道联合编码（压缩属于信源编码）。虽然香农指出，在AWGN信道下，对信源编码和信道码进行分离设计也能达到最优[58]，但其前提是无限的信源编码维数、无限的信道编码分组长度以及无限的复杂度和时延。因此，人们从一些实际的条件出发，对信源信道联合编码问题开展了大量的工作。

信源信道联合编码方面的工作主要分为信源优化的信道编码、信道优化的信源编码以及迭代算法这样几类。在信源优化的信道编码中，信源编码按信道无噪声设计，再针对这个信源编码设计信道编码，以使端到端的失真最小。UEP编码就是这样的例子，它使不同的子信道编码的误比特率匹配于不同信源编码比特的优先级。文献[54]及[59~60]对图像传输采用了信源最优的信道编码，它们将压缩图像通过速率匹配的凿孔卷积码（rate-compatibel punctured convolutional channel codes，RCPC）传输。文献[61]针对AWGN信道和瑞利衰落信道，研究了RCPC或MQAM同子带线性预测语音编码的匹配问题。在信源最优的调制中，信源编码按无噪声的信道条件设计，再设计调制使端到端的失真最小。例如在文献[62]中，矢量量化器（vector quantizer，VQ）的压缩结果通过多载波调制传输，而多载波的调制再通过子载波功率分配来实现不同的信源比特的不同错误保护。

274

在信道优化的信源编码中，信道码独立于信源进行设计，再按信道编码的误码率优化信源编码的设计。这种例子比如有信道优化的矢量量化器（channel-optimized vector quantizer, COVQ）和标量量化器[63~64]。信道优化的信源编码与信源优化的信道编码或调制可以结合起来进行迭代设计，例如文献[65]结合了COVQ和多载波调制，文献[66]结合了COVQ和RCPC，文献[67~68]结合了格码调制和格码量化（trellis coded quantization)，其中格码量化是一种借鉴了格码调制思想的信源编码。这些研究结果表明，对信源编码和信道编码进行联合设计能带来可观的性能增益。许多先进的信道编码，如Turbo码和LDPC码，还没有同信源编码一起进行最优的联合设计。针对不同的应用，信源信道联合编码方面还有大量工作要做。

习题

8.1 已知(3, 1)线性分组码的每个码字都包括两个数据比特和一个校验比特。

(a) 写出该码中的所有码字。

(b) 计算该码的最小距离。

8.2　已知 (7, 4) 码的生成矩阵为

$$G=\begin{bmatrix}0&1&0&1&1&0&0\\1&0&1&0&1&0&0\\0&1&1&0&0&1&0\\1&1&0&0&0&0&1\end{bmatrix}$$

(a) 写出该码的所有码字。

(b) 求该码的最小码距。

(c) 写出该码的校验矩阵。

(d) 求接收向量$\boldsymbol{R}=[1101011]$的伴随式。

(e) 设发送的信息比特全为零，发生错误图样$\boldsymbol{e}$后的接收码字依然是合法码字，但不是全零码字。试找出具有最小码重的这种错误图案。

(f) 利用行列变换将$\boldsymbol{G}$转化为系统形式，写出对应的校验矩阵。画出用移位寄存器实现该系统码的框图。

275

8.3　所有汉明码的最小码距都是3，请问汉明码的纠错能力和检错能力是多少？

8.4　已知(15, 11)汉明码的生成多项式为$g(X)=1+X+X^4$。试判断多项式$c_1(X)=1+X+X^3+X^7$和$c_2(X)=1+X^3+X^5+X^6$代表的码字是不是该码的合法码字，并将其写成对应于系统码的形式$p(X)+X^{n-k}u(X)$。

8.5　已知(7, 4)循环汉明码的生成多项式为$g(X)=1+X^2+X^3$。

(a) 写出该码系统码形式的生成矩阵。

(b) 写出该码的校验矩阵。

(c) 假设发送码字$\boldsymbol{C}=[1011010]$对应的接收码字为$\boldsymbol{R}=[1010011]$，求其伴随式。

(d) 对于发送码字$\boldsymbol{C}=[1011010]$，写出所有能使伴随式为0的接收码字。

8.6　码长为n的汉明码的码重分布为

$$N(x)=\sum_{i=0}^{n}N_ix^i=\frac{1}{n+1}[(1+x)^n+n(1+x)^{0.5(n-1)}(1-x)^{0.5(n+1)}]$$

其中N_i是码重为i的码字数目。

(a) 利用该公式计算(7, 4)汉明码的码重分布。

(b) 假设BPSK调制，信噪比为10dB。请利用 (a) 中得到的码重分布求出式(8-38)的错误率联合界，并同松散界式 (8-39) 进行比较。

8.7　假设编码比特采用8PSK调制在信噪比为10dB的AWGN信道中传输，请写出$m=7$的汉明码的码字错误率的联合界。如果一个码字错误对应只发生一个比特错误，试计算该码的误比特率，并与不采用编码直接用8PSK传输时的误比特率进行比较。

8.8　某 (24, 12) 线性分组码的最小码距是$d_{\min}=8$，编码比特通过QPSK调制传输。请利用码字错误率的联合界，画出$0\leqslant E_b/N_0\leqslant 20$dB范围内$P_b$与比特信噪比$\gamma_b$的关系曲线。请问，当$E_b/N_0$处于什么范围时，该码是否会表现出负编码增益？

8.9　已知$\gamma_b=15$dB，请用式 (8-47) 计算 (7, 4) 汉明码软判决译码时编码增益的

近似值。

8.10 某（24, 12）线性分组码的最小码距是$d_{min}=8$，编码比特通过BPSK传输。请利用式（8-36）和式（8-44）画出$0\leq\gamma_b\leq10$dB范围内硬判决译码和软判决译码的码字错误率曲线。276高信噪比时，这两种译码的编码增益相差多少？

8.11 假设编码符号以BPSK调制在信噪比为10dB的AWGN信道中传输，请利用式（8-35）和式（8-36）计算扩展戈雷码的码字错误率的上界和下界。

8.12 某RS码的$k=3$、$K=4$，映射为8PSK调制。试求该码可纠正的错误数及最小码距。假设8PSK调制的误码率为10^{-3}，计算RS译码后的比特错误率。

8.13 对于图8-7所描绘的格图，已知初始状态$S=00$，信息比特序列为$U=[0110101101]$，请写出相应的状态序列和编码输出。

8.14 考虑图8-24所示的卷积码编码器。

（a）画出该码的格图。

（b）若信道的符号错误率为$p=10^{-3}$，接收序列为$\boldsymbol{R}=[001010001]$，请计算全零路径的路径度量。

（c）找出一条与全零路径汉明距离最小的路径，假设$\boldsymbol{R}$和p都与（b）小问一样，计算出这条路径的路径度量。

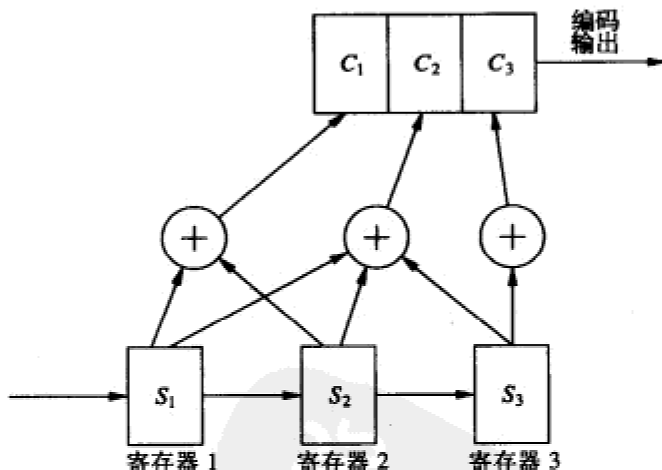


图8-24 题8.14和题8.15中的卷积码编码器

8.15 考虑图8-24中的卷积编码器。

（a）画出该卷积码的状态图。

（b）求转移函数$T(D, N, J)$。

（c）找出所有非全零路径与全零路径之间的最小距离。

（d）假设采用BPSK调制和软判决译码，$\gamma_b=10$dB。请利用式（8-72）计算该码的误比特率上界。

（e）假设采用BPSK调制和硬判决译码，$\gamma_b=10$dB，请利用式（8-76）计算该码的误比特率上界。软判决译码相对于硬判决译码有多少增益？277

8.16　某网格码调制系统如图8-25所示，其中16-QAM的子集划分方式见图8-18。

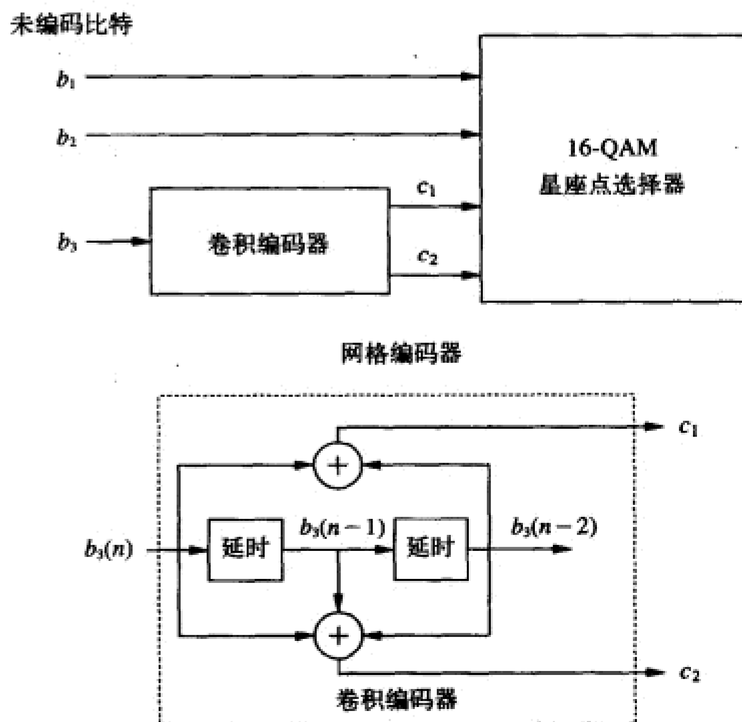


图8-25　题8.16中的16-QAM网格编码器

(a) 假设错误主要发生在子集内，请计算此格码调制相对于未编码的8PSK的编码增益。已知16QAM的d_0是0.632，8PSK的d_0是0.765；

(b) 画出该卷积编码器的格图，并根据如下Ungerboeck提出的经验规则将子集分配给格图转移：(i) 如果每个调制符号间隔内有k个比特需要编码，那么必须允许从每个状态到下一个状态中有$2k$次可能的转移。(ii) 在一对状态之间可以发生超过一次的转移。(iii) 所有波形的出现机会应该有相同，并有一定的规则性和对称性。(iv) 从同一状态发起的转移可以分配给A子集中的波形或者B子集中的波形，但不能都给A子集或都给B子集。(v) 进入同一状态的转移可以分配给A子集中的波形或者B子集中的波形，但不能都给A子集或都给B子集。(vi) 并行转移只能分配给一个子集。

(c) 相对于全零比特流产生的路径，找出格图中具有最小距离的错误事件。

(d) 假设 (c) 中的结果就是格图的最小距离错误事件，该码的d_{min}是多少？

278

8.17　已知信道的相干时间为$T_c=10$ms，编码比特的传输速率为$R_s=50$kbit/s。对于卷积交织器，欲使相邻编码比特之间的衰落独立，求平均交织时延及所需要的存储量。

8.18　求深交织（$dT_s>>T_c$）的戈雷码（23，12）的误比特率P_b上界，并与未编码的情况进行比较。假设为瑞利衰落信道、BPSK调制、软判决译码、编码比特的平均$E_c/N_0=15$dB。

8.19　假设瑞利衰落信道、BPSK调制、平均信噪比是10dB、多普勒频移为80Hz、信道中的传输速率为30kbit/s。假设当$P_b(\gamma)\geq10^{-2}$时信道就会发生比特错误。试设计一个交织器及

相应的能基本纠正所有的比特错误（n, k）分组码。要求交织时延不能大于50ms。设计应包括交织器的维数、分组码的类型以及n和k的数值。

8.20　假设图8-21的多电平编码器中信息比特有三种不同的错误保护级别（$M=3$），三个编码器的输出映射为8PSK调制，并以10dB的信噪比传输。假设与第i个比特流b_i对应的码C_i是参数为m_i的汉明码，$m_1=2$、$m_2=3$、$m_3=4$。

（a）假设每个汉明码C_i都是独立的硬判决译码，求各自的错误概率。

（b）若8PSK的码元间隔为$T_s=10\mu s$，那么这三个比特流的数据速率各是多少？

（c）若对该UEP码进行最大似然译码，码本大小是多少？

8.21　利用汉明码或者戈雷码设计一个两级的UEP码，使得信噪比为10dB时，低优先级比特的错误率为$P_b=10^{-3}$，高优先级比特的错误率为$P_b=10^{-6}$。

参考文献

[1] J. G. Proakis, *Digital Communications,* 4th ed., McGraw-Hill, New York, 2001.

[2] B. Sklar, *Digital Communications – Fundamentals and Applications,* Prentice-Hall, Englewood Cliffs, NJ, 1988.

[3] D. G. Wilson, *Digital Modulation and Coding,* Prentice-Hall, Englewood Cliffs, NJ, 1996.

[4] S. Lin and J. D. J. Costello, *Error Control Coding,* 2nd ed., Prentice-Hall, Englewood Cliffs, NJ, 2004.

[5] A. J. Viterbi, "Error bounds for convolutional codes and asymptotically optimum decoding algorithm," *IEEE Trans. Inform. Theory,* pp. 260–9, April 1967.

[6] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: Turbo-codes," *Proc. IEEE Internat. Conf. Commun.*, pp. 54–8, May 1993.

[7] C. Berrou and A. Glavieux, "Near optimum error correcting coding and decoding: Turbo-codes," *IEEE Trans. Commun.*, pp. 1261–71, October 1996.

[8] *IEEE Trans. Inform. Theory,* Special Issue on Codes and Graphs and Iterative Algorithms, February 2001.

[9] S. B. Wicker and S. Kim, *Codes, Graphs, and Iterative Decoding,* Kluwer, Boston, 2002.

[10] R. G. Gallager, "Low-density parity-check codes," *IRE Trans. Inform. Theory,* pp. 21–8, January 1962.

279

[11] C. Heegard and S. B. Wicker, *Turbo Coding,* Kluwer, Boston, 1999.

[12] M. E. Buckley and S. B. Wicker, "The design and performance of a neural network for predicting decoder error in turbo-coded ARQ protocols," *IEEE Trans. Commun.*, pp. 566–76, April 2000.

[13] S. Benedetto, D. Divsalar, G. Montorsi, and F. Pollara, "Serial concatenation of interleaved codes: Performance analysis, design and iterative decoding," *IEEE Trans. Inform. Theory,* pp. 909–26, May 1998.

[14] H. Jin and R. J. McEliece, "Coding theorems for turbo code ensembles," *IEEE Trans. Inform. Theory,* pp. 1451–61, June 2002.

[15] I. Sasan and S. Shamai, "Improved upper bounds on the ML decoding error probability of parallel and serial concatenated turbo codes via their ensemble distance spectrum," *IEEE Trans. Inform. Theory,* pp. 24–47, January 2000.

[16] D. J. C. MacKay and R. M. Neal, "Near Shannon limit performance of low density parity check codes," *Elec. Lett.*, p. 1645, August 1996.

[17] N. Wiberg, N.-A. Loeliger, and R. Kotter, "Codes and iterative decoding on general graphs," *Euro. Trans. Telecommun.*, pp. 513–25, June 1995.

[18] T. Richardson, A. Shokrollahi, and R. Urbanke, "Design of capacity-approaching irregular low-density parity-check codes," *IEEE Trans. Inform. Theory,* pp. 619–37, February 2001.
[19] R. McEliece, D. J. C. MacKay, and J.-F. Cheng, "Turbo decoding as an instance of Pearl's 'belief propagation' algorithm," *IEEE J. Sel. Areas Commun.,* pp. 140–52, February 1998.
[20] F. R. Kschischang and D. Frey, "Iterative decoding of compound codes by probability propagation in graphical models," *IEEE J. Sel. Areas Commun.,* pp. 219–30, February 1998.
[21] T. Richardson and R. Urbanke, "The capacity of low-density parity-check codes under message passing decoding," *IEEE Trans. Inform. Theory,* pp. 599–618, February 2001.
[22] S.-Y. Chung, G. D. Forney, T. Richardson, and R. Urbanke, "On the design of low-density parity-check codes within 0.0045 dB of the Shannon limit," *IEEE Commun. Lett.,* pp. 58–60, February 2001.
[23] M. Fossorier, "Iterative reliability-based decoding of low-density parity check codes," *IEEE J. Sel. Areas Commun.,* pp. 908–17, May 2001.
[24] M. C. Davey and D. MacKay, "Low density parity-check codes over GF(q)," *IEEE Commun. Lett.,* pp. 165–7, June 1998.
[25] J. Hou, P. Siegel, L. Milstein, and H. D. Pfister, "Capacity-approaching bandwidth efficient coded modulation schemes based on low-density parity-check codes," *IEEE Trans. Inform. Theory,* pp. 2141–55, September 2003.
[26] G. Ungerboeck, "Channel coding with multi-level/phase signals," *IEEE Trans. Inform. Theory,* pp. 55–67, January 1982.
[27] G. D. Forney, "Coset codes, I: Introduction and geometrical classification, and II: Binary lattices and related codes," *IEEE Trans. Inform. Theory,* pp. 1123–87, September 1988.
[28] G. D. Forney, Jr., R. G. Gallager, G. R. Lang, F. M. Longstaff, and S. U. Quereshi, "Efficient modulation for band-limited channels," *IEEE J. Sel. Areas Commun.,* pp. 632–47, September 1984.
[29] S. Benedetto, D. Divsalar, G. Montorsi, and F. Pollara, "Parallel concatenated trellis coded modulation," *Proc. IEEE Internat. Conf. Commun.,* pp. 974–8, June 1996.
[30] C. Fragouli and R. D. Wesel, "Turbo-encoder design for symbol-interleaved parallel concatenated trellis-coded modulation," *IEEE Trans. Commun.,* pp. 425–35, March 2001.
[31] P. Robertson and T. Worz, "Bandwidth-efficient turbo trellis-coded modulation using punctured component codes," *IEEE J. Sel. Areas Commun.,* pp. 206–18, February 1998.
[32] G. T. Irvine and P. J. Mclane, "Symbol-aided plus decision-directed reception for PSK TCM modulation on shadowed mobile satellite fading channels," *IEEE J. Sel. Areas Commun.,* pp. 1289–99, October 1992.
[33] D. Subasinghe-Dias and K. Feher, "A coded 16-QAM scheme for fast fading mobile radio channels," *IEEE Trans. Commun.,* pp. 1906–16, February–April 1995.
[34] P. Y. Kam and H. M. Ching, "Sequence estimation over the slow nonselective Rayleigh fading channel with diversity reception and its application to Viterbi decoding," *IEEE J. Sel. Areas Commun.,* pp. 562–70, April 1992.
[35] D. Makrakis, P. T. Mathiopoulos, and D. P. Bouras, "Optimal decoding of coded PSK and QAM signals in correlated fast fading channels and AWGN – A combined envelope, multiple differential and coherent detection approach," *IEEE Trans. Commun.,* pp. 63–75, January 1994.
[36] M. J. Gertsman and J. H. Lodge, "Symbol-by-symbol MAP demodulation of CPM and PSK signals on Rayleigh flat-fading channels," *IEEE Trans. Commun.,* pp. 788–99, July 1997.
[37] H. Kong and E. Shwedyk, "Sequence detection and channel state estimation over finite state Markov channels," *IEEE Trans. Veh. Tech.,* pp. 833–9, May 1999.
[38] G. M. Vitetta and D. P. Taylor, "Maximum-likelihood decoding of uncoded and coded PSK signal sequences transmitted over Rayleigh flat-fading channels," *IEEE Trans. Commun.,* pp. 2750–8, November 1995.

280

PDG

[39] L. Li and A. J. Goldsmith, "Low-complexity maximum-likelihood detection of coded signals sent over finite-state Markov channels," *IEEE Trans. Commun.*, pp. 524–31, April 2002.

[40] B. Vucetic and J. Yuan, *Turbo Codes: Principles and Applications*, Kluwer, Dordrecht, 2000.

[41] E. K. Hall and S. G. Wilson, "Design and analysis of turbo codes on Rayleigh fading channels," *IEEE J. Sel. Areas Commun.*, pp. 160–74, February 1998.

[42] C. Komninakis and R. D. Wesel, "Joint iterative channel estimation and decoding in flat correlated Rayleigh fading," *IEEE J. Sel. Areas Commun.* pp. 1706–17, September 2001.

[43] J. Hou, P. H. Siegel, and L. B. Milstein, "Performance analysis and code optimization of low-density parity-check codes on Rayleigh fading channels," *IEEE J. Sel. Areas Commun.*, pp. 924–34, May 2001.

[44] A. J. Goldsmith and P. P. Varaiya, "Capacity, mutual information, and coding for finite-state Markov channels," *IEEE Trans. Inform. Theory*, pp. 868–86, May 1996.

[45] G. D. Forney, "Burst error correcting codes for the classic bursty channel," *IEEE Trans. Commun. Tech.*, pp. 772–81, October 1971.

[46] J. L. Ramsey, "Realization of optimum interleavers," *IEEE Trans. Inform. Theory*, pp. 338–45, 1970.

[47] C.-E. W. Sundberg and N. Seshadri, "Coded modulation for fading channels – An overview," *Euro. Trans. Telecommun.*, pp. 309–24, May/June 1993.

[48] L.-F. Wei, "Coded M-DPSK with built-in time diversity for fading channels," *IEEE Trans. Inform. Theory*, pp. 1820–39, November 1993.

[49] S. H. Jamali and T. Le-Ngoc, *Coded-Modulation Techniques for Fading Channels*, Kluwer, New York, 1994.

[50] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. Inform. Theory*, pp. 927–46, May 1998.

[51] E. Zehavi, "8-PSK trellis codes for a Rayleigh channel," *IEEE Trans. Commun.*, pp. 873–84, May 1992.

[52] H. Imai and S. Hirakawa, "A new multilevel coding method using error correcting codes," *IEEE Trans. Inform. Theory*, pp. 371–7, May 1977.

[53] R. V. Cox, J. Hagenauer, N. Seshadri, and C.-E. W. Sundberg, "Variable rate sub-band speech coding and matched convolutional channel coding for mobile radio channels," *IEEE Trans. Signal Proc.*, pp. 1717–31, August 1991.

[54] J. W. Modestino and D. G. Daut, "Combined source-channel coding of images," *IEEE Trans. Commun.*, pp. 1644–59, November 1979.

[55] A. R. Calderbank and N. Seshadri, "Multilevel codes for unequal error protection," *IEEE Trans. Inform. Theory*, pp. 1234–48, July 1993.

[56] L.-F. Wei, "Coded modulation with unequal error protection," *IEEE Trans. Commun.*, pp. 1439–49, October 1993.

[57] N. Seshadri and C.-E. W. Sundberg, "Multilevel trellis coded modulations for the Rayleigh fading channel," *IEEE Trans. Commun.*, pp. 1300–10, September 1993.

281

[58] C. E. Shannon, "Coding theorems for a discrete source with a fidelity criterion," *IRE National Convention Record*, part 4, pp. 142–63, 1959.

[59] N. Tanabe and N. Farvardin, "Subband image coding using entropy-coded quantization over noisy channels," *IEEE J. Sel. Areas Commun.*, pp. 926–43, June 1992.

[60] H. Jafarkhani, P. Ligdas, and N. Farvardin, "Adaptive rate allocation in a joint source/channel coding framework for wireless channels," *Proc. IEEE Veh. Tech. Conf.*, pp. 492–6, April 1996.

[61] W. C. Wong, R. Steele, and C.-E. W. Sundberg, *Source-Matched Mobile Communications*, Pentech and IEEE Press, London and New York, 1995.

[62] K.-P. Ho and J. M. Kahn, "Transmission of analog signals using multicarrier modulation: A combined source-channel coding approach," *IEEE Trans. Commun.*, pp. 1432–43, November 1996.

[63] N. Farvardin and V. Vaishampayan, "On the performance and complexity of channel-optimized vector quantizers," *IEEE Trans. Inform. Theory*, pp. 155–60, January 1991.

[64] N. Farvardin and V. Vaishampayan, "Optimal quantizer design for noisy channels: An approach to combined source-channel coding," *IEEE Trans. Inform. Theory*, pp. 827–38, November 1987.

[65] K.-P. Ho and J. M. Kahn, "Combined source-channel coding using channel-optimized quantizer and multicarrier modulation," *Proc. IEEE Internat. Conf. Commun.*, pp. 1323–7, June 1996.

[66] A. J. Goldsmith and M. Effros, "Joint design of fixed-rate source codes and multiresolution channel codes," *IEEE Trans. Commun.*, pp. 1301–12, October 1998.

[67] E. Ayanoglu and R. M. Gray, "The design of joint source and channel trellis waveform coders," *IEEE Trans. Inform. Theory*, pp. 855–65, November 1987.

[68] T. R. Fischer and M. W. Marcellin, "Joint trellis coded quantization/modulation," *IEEE Trans. Commun.*, pp. 172–6, February 1991.

282



第9章　自适应调制编码

在时变信道中，自适应调制编码可以增强传输的可靠性并能提高频带利用率。自适应调制编码的基本思想是，在接收端估计信道，然后将估计结果反馈到发送端，发送端根据信道的特性调整发送方式。非自适应的系统是按照最坏情况设计的，为了在信道条件变得很差时也能维持可接受的性能，需要留出一个固定的链路余量。瑞利衰落中，由此引起的信号功率损失可能高达30dB，所以按最坏情况进行的设计必然导致非常低效的信道利用。自适应方式在信道条件好的时候提高传速率或减小发送功率，在信道条件差的时候降低速率或增大功率，从而可以提高平均吞吐量，降低所需的发射功率，或降低平均误码率。我们在4.2.4节中已经导出了平衰落信道中能达到香农容量的最佳自适应传输方案，本章将介绍一些更为实际的自适应调制和编码技术，它们能使平均的频谱利用率最大化，同时能保证给定的平均误码率或瞬时误码率要求。本章所介绍的自适应概念同样适用于MIMO信道、带均衡的频率选择性信道、OFDM或CDMA系统以及蜂窝系统。自适应技术在这些系统中的应用将在后续的章节中进行介绍。

人们对自适应传输技术的研究始于20世纪60年代末70年代初[1~2]，不过研究兴趣没有持续多长时间，原因也许是因为硬件方面的限制，缺少好的信道估计技术或者是因为当时人们主要关注的是点到点无线链路，没有向发送端的反馈。随着技术的发展，这些限制逐渐消失，发展到3G无线系统时，自适应技术的研究又开始活跃了起来[3~12]。包括GSM、CDMA蜂窝系统及无线局域网在内的许多无线系统都已经采用了自适应传输技术[13~16]。

283

自适应发送技术的应用存在一些实际限制。首先，它要求发送端和接收端之间存在反馈通路，这在某些系统中是不大可能的。其次，如果信道变化的速度快于信道估计及反馈的速度，自适应技术的性能将会很差。无线信道中一般存在着不同时间刻度的时变性，例如多径衰落变化非常快，阴影衰落变化很慢。许多系统只能根据慢衰落进行自适应发送，在这种情况下，还需要一些对抗平衰落的措施来解决多径的影响。硬件因素会制约发送端改变传输速率和功率的快慢，使自适应技术能带来的增益受到限制。最后，自适应技术经常会根据信道的条件改变数据传输速率，在平均功率约束下，为了最大化频谱效率，信道条件差时数据速率可能会调到很小甚至为零。这对于有硬性时延要求的固定速率业务，如语音和视频来说，性能将受到严重影响。因此对于有时延约束的应用，自适应发送技术应该针对固定的传输速率来优化中断率[17]。

9.1　自适应传输系统

本节给出的自适应发送系统的模型和4.2.1节研究平衰落信道容量时的模型一样。假设线性调制，自适应调整发生在码元间隔T_s的整倍数处。同时假设采用理想奈奎斯特成形脉冲$\mathrm{sinc}[t/T_s]$，调制信号带宽为$B=1/T_s$。考虑离散时间的平衰落信道模型，信道模

PDG

型中的采样间隔等于码元间隔T_s。信道的时变增益$\sqrt{g[i]}$是平稳遍历过程，其概率密度函数为$p(g)$。信道噪声$n[i]$是加性白高斯噪声，其功率谱密度为$N_0/2$。令$\overline{P}$表示平均发送功率、$B=1/T_s$表示接收信号的带宽、$\overline{g}$表示平均信道增益，则瞬时接收信噪比为$\gamma[i]=\overline{P}g[i]/N_0B$，$0\leqslant\gamma[i]<\infty$，平均信噪比为$\overline{\gamma}=\overline{P}\,\overline{g}/N_0B$。由于$g[i]$是平稳的，所以$\gamma[i]$的概率密度函数与$i$无关，记其为$p(\gamma)$。

在自适应传输中，我们在时刻i估计功率增益或接收信噪比，再相应调整调制和编码的参数。可调整的参数一般包括数据速率$R[i]$、发送功率$P[i]$和编码参数$C[i]$。M进制调制中，数据速率$R[i]=\log_2M[i]/T_s=B\log_2M[i]$bit/s，频谱利用率为$R[i]/B=\log_2M[i]$bit/s/Hz。记估计的信噪比为$\hat{\gamma}[i]=\overline{P}\hat{g}[i]/N_0B$，其中$\hat{g}[i]$是功率增益的估计值。假设根据$\hat{\gamma}[i]$来调整发送功率，则记$i$时刻的发送功率为$P(\hat{\gamma}[i])=P[i]$，$i$时刻的接收功率便为$\gamma[i]\ P(\hat{\gamma}[i])/\overline{P}$。类似的，根据$\hat{\gamma}[i]$来调整的数据速率是$R(\hat{\gamma}[i])=R[i]$，编码参数是$C(\hat{\gamma}[i])=C[i]$。在不致混淆的情形下，以下将省略$\gamma$、$P(\gamma)$、$R(\gamma)$、$C(\gamma)$中的时间参数$i$。

系统模型如图9-1所示。假设经过时延i_e后，接收端得到了i时刻信道的功率增益$g[i]$的估计值$\hat{g}[i]$，再加上反馈时延i_f，发送端在经过时延$i_d=i_e+i_f$后得到了这个信息。发送端依据这个信道信息来实现对信道变化的自适应。具体做法很多，可以直接把$\hat{g}[i-i_d]$当作真实值来用，也可以考虑$\hat{g}[i-i_d]$中存在的估计误差和时延因素，9.3.7节将对此进行详细讨论。我们假设反馈是理想的，在反馈信道采用了高性能纠错编码及检错重传机制时，这是一个合理的假设。

284

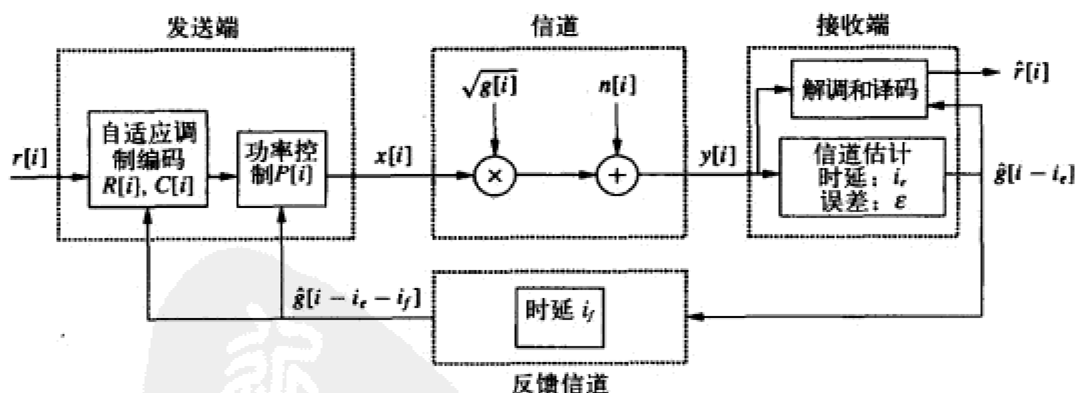


图9-1 系统模型

信道变化的速度决定了发送端调整传输参数的速度，也同样会影响$g[i]$的估计误差。当信道增益同时包含快衰落和慢衰落时，如果$g[i]$变化足够慢，发送端将能同时适应这两种衰落，否则只能适应慢衰落。比如当$g[i]$代表阴影衰落和多径衰落的总效果时，低速情况下阴影衰落是常数，多径衰落也很慢，这样就可以估计出信道信息并反馈到发送端，估计误差和反馈时延不会使性能明显下降。而在高速时，系统无法做到有效估计并反馈多径衰落的信息，此时系统只是对阴影变化进行适应，不再适应多径衰落，这种情况下需要对快衰落进行平均。我们将在9.5节中讨论结合快衰落和慢衰落的自适应技术。

9.2 自适应技术

依据信道增益γ，发送端可控制的参数有很多。本节讨论针对最常见的一些参数进行调整的自适应技术，包括数据速率、功率、编码、错误率以及它们的结合。

9.2.1 速率控制技术

速率控制技术根据信道增益γ来控制发送的数据速率$R[\gamma]$。具体实现方式可以是固定码元速率$R_s=1/T_s$，改变调制方式或星座大小，也可以是固定调制方式而改变符号速率。改变符号速率会引起带宽的改变，这在实现上有一些困难，同时改变符号速率也使频带共享的问题复杂化。而固定码元速率，改变调制方式或星座大小的方法容易实现，现有的许多系统，包括GSM、IS-136 EDGE以及802.11a无线局域网等许多系统就是根据信道质量来改变调制和编码方案的[15]。

285

在速率控制技术中，与传输速率相对应的调制参数在一帧内是固定的，帧大小是一个设计参数。每帧中还可能包括用于信道估计的导频符号和其他控制信息。

当所采用的调制类型或星座大小是离散集合时，需要把每个γ值映射到一种调制方式上。映射时一般要求每一种调制方式的误比特率都能低于某个给定值。这一点如下面的例题所示，后面几节将要讲到的一些具体的自适应调制技术也是如此。

例9.1：某自适应调制系统使用QPSK和8PSK两种调制方式，目标P_b近似为10^{-3}。如果这两种调制都不能达到目标P_b，则不发送数据。求不传输、QPSK、8-PSK这三种传输方式对应的γ值范围，以及系统的平均频谱利用率。假设瑞利衰落，$\overline{\gamma}=20\text{dB}$。

解：对QPSK和8-PSK都有$\gamma=\gamma_s$。由6.1节可知，对QPSK，$P_b\approx Q(\sqrt{\gamma})$，对8-PSK，$P_b\approx 0.666Q\left(\sqrt{2\gamma}\sin(\pi/8)\right)$。8-PSK在$\gamma>14.79\text{dB}$时$P_b<10^{-3}$，所以使用8-PSK的信噪比范围是$\gamma>14.79\text{dB}$。QPSK在$\gamma>10.35\text{dB}$时$P_b<10^{-3}$，所以使用QPSK的信噪比范围是$14.79\text{dB}\geqslant\gamma>10.35\text{dB}$。$\gamma\leqslant10.35\text{dB}$时不传数据。

我们根据不同传输方式所占时间的比例来计算平均速率。$\gamma>14.79\text{dB}=30.1$时使用8-PSK，其频带利用率为$R[\gamma]/B=\log_2 8=3\text{bit/s/Hz}$，在$\overline{\gamma}=20\text{dB}$的瑞利衰落下，此传输方式的传输时间所占的比例为$P_8=\int_{30.1}^{\infty}\frac{1}{100}e^{-\gamma/100}d\gamma=0.74$。$14.79\text{dB}\geqslant\gamma>10.35\text{dB}$时使用QPSK，式中$10.35\text{dB}=10.85$，所以$R[\gamma]=\log_2 4=2\text{bit/s/Hz}$，传输时间所占的比例为$P_4=\int_{10.85}^{30.1}\frac{1}{100}e^{-\gamma/100}d\gamma=0.157$。在剩下的0.103份时间内，没有数据传输。因此平均的频带利用率为$0.74\times3+0.157\times2+0.103\times0=2.534\text{bit/s/Hz}$。

如果在$\gamma\leqslant10.35\text{dB}$时继续传输，就会有0.1的中断率。此时也可以考虑使用BPSK这样的一维调制，或者用带纠错编码的QPSK来保证目标P_b。如果是采用纠错编码，那么$\gamma\leqslant10.35\text{dB}$时的频带利用率将小于2bit/s/Hz，但比不传输时的0大。9.2.4节中将介绍基于编码调整的自适应技术。

9.2.2 功率控制技术

单纯的功率控制一般只是为了补偿衰落引起的信噪比变化，其目标是维持一个固定的误比特率或者说是维持一个固定的接收信噪比。这种功率自适应方案实际是反转了信道的衰落，使得调制器和解调器之间的信道如同AWGN信道[1]。信道反转的功率自适应方案为：

$$\frac{P(\gamma)}{\bar{P}}=\frac{\sigma}{\gamma} \tag{9-1}$$

286

式中σ是固定的接收信噪比。在平均功率受限于$\bar{P}$的条件下，我们有

$$\int \frac{P(\gamma)}{\bar{P}}p(\gamma)\,\mathrm{d}\gamma=\int \frac{\sigma}{\gamma}p(\gamma)\,\mathrm{d}\gamma=1 \tag{9-2}$$

由此可解得$\sigma=1/E[1/\gamma]$。σ与$p(\gamma)$有关，$p(\gamma)$又通过$\bar{\gamma}$与平均发送功率$\bar{P}$有关。因此，给定平均功率$\bar{P}$时，如果σ的值大于$1/E[1/\gamma]$，就不能达到目标BER。注意瑞利衰落中γ是指数分布，$E[1/\gamma]=\infty$，因此对于瑞利衰落，无论P_b是多少，用信道反转方案都不能达到目标。

也可以只在衰落高于给定的截止门限γ_0时，才进行信道反转，这就是截断式信道反转的功率自适应方案：

$$\frac{P(\gamma)}{\bar{P}}=\begin{cases}\sigma/\gamma & \gamma\geqslant\gamma_0\\ 0 & \gamma<\gamma_0\end{cases} \tag{9-3}$$

门限γ_0可根据要求的中断率$P_{out}=p(\gamma<\gamma_0)$选取，也可根据门限之上的目标误码率来选取，此时门限值与目标误码率和$p(\gamma)$都有关。因为只有当$\gamma\geqslant\gamma_0$时才使用该信道，所以在给定平均功率$\bar{P}$的条件下，有$\sigma=1/E_{\gamma_0}[1/\gamma]$。其中

$$E_{\gamma_0}\left[\frac{1}{\gamma}\right]\triangleq\int_{\gamma_0}^{\infty}\frac{1}{\gamma}p(\gamma)\,\mathrm{d}\gamma \tag{9-4}$$

例9.2：假设瑞利衰落信道，$\bar{\gamma}=10\text{dB}$，BPSK调制。如要达到正常通信时的$P_b=10^{-3}$，求相应的功率自适应方案，并求相应的中断率。

解：功率自适应方案为截断式信道反转，所以只需求出σ和γ_0。BPSK在信噪比为$\sigma=4.77$时$P_b=Q\left(\sqrt{2\sigma}\right)=10^{-3}$。令$\sigma=1/E_{\gamma_0}[1/\gamma]$，求解$\gamma_0$。这个问题只能用数值方法求解，结果是$\gamma_0=0.7423$。相应的中断率为$P_{out}=p(\gamma<\gamma_0)=1-\mathrm{e}^{-\gamma_0/10}=0.379$。中断率高的原因是这个信道条件相对于所要求$P_b=10^{-3}$来说比较差。

9.2.3 错误率控制技术

同样也可以在平均BER给定为$\bar{P}_b$的条件下，控制瞬时BER。从6.3.2节可知，衰落信道中的瞬时错误率随接收信噪比的变化而变化，平均的误比特率为$\bar{P}_b=\int P_b(\gamma)p(\gamma)\mathrm{d}\gamma$。瞬时错误率的这种变化当然不是自适应技术，因为发送端并没有根据γ而调整什么。在自

1. 4.2.4节分析衰落信道容量时，已经讨论了信道反转和截断式信道反转。

适应调制中，控制瞬时错误率的自适应技术其实是通过其他形式的自适应技术实现的，比如结星座大小调整或调制方式调整。Hayes在文献[1]中最早提出了同时控制数据速率和错误率的技术。这种技术能降低发送能量，对应目标平均BER为10^{-4}的情形，可节省4dB的功率。

287

9.2.4 编码控制技术

自适应编码通过不同的信道编码对传输比特提供不同的编码增益。比如，当γ比较小时使用纠错能力强的编码，而当γ比较大时使用纠错能力弱一些的编码或者不进行编码。自适应编码的一种实现方式是设计多个不同纠错能力的编码，不同信道条件时连接到不同的编码器。这种方法需要在码字长度或编码的约束长度内信道大致不变[7]。出于复杂度或峰均功率比等方面的考虑，有时我们必须要固定调制方式，自适应编码技术非常适合于这种情况。

实现自适应编码的另一种方法是采用码率兼容凿孔卷积码（RCPC）[18]。RCPC码是基于同一个编码器和译码器的一族不同码率的卷积码。凿孔的意思就是不发送编码结果中的某些比特。通过凿孔就能控制码的纠错能力。RCPC码具有码率兼容性，高码率码（纠错能力弱）的编码比特也出现在所有低码率码（纠错能力强）中。增加编码比特的个数就能增加码的纠错能力。RCPC的码率兼容性使其易于调整差错保护能力，不同的码率都是同一个编码器和译码器，只要发送端采用适当的凿孔就可以达到所需要的差错保护能力。RCPC的译码器按最低码率的码用维特比算法进行译码。对于高码率，译码器在计算分支度量时考虑到了发送端凿掉的比特。凿孔是一种简单有效的自适应编码技术，GSM和IS-136 EDGE中的自适应编码就是以此为基础的。

上述的自适应编码技术可以用于固定调制方式，也可以结合自适应调制。当调制方式必须固定时，只能靠自适应编码来适应信道变化[6-7]。本章主要考虑可以改变调制方式的自适应系统，所以自适应编码本身的问题就不再进一步讨论。

9.2.5 混合技术

混合的自适应技术可以同时控制多个发送参数，包括速率、功率、编码和瞬时错误率等。它通过不同参数的联合优化来达到预期的性能。自适应速率经常与自适应功率结合在一起以优化频谱效率，本章后面几节将把这种联合优化应用到不同的调制类型中。有大量文献在研究自适应调制和自适应编码，现有的蜂窝系统和无线局域网也已经采用了这种技术[13],[15]。

288

9.3 速率可变功率可变的MQAM

前一节讨论了自适应调制编码的一般方法。本节考虑一种具体的自适应调制，它在满足瞬时目标P_b的前提下，通过控制MQAM的速率和功率来优化频谱利用率。研究这种自适应调制的原因一是可以使我们深入理解自适应调制的优点，二是因为对它而言最优的功率和速率控制方案同样也是能达到容量的功率和速率控制方案。我们将看到，在

平衰落信道中，这种自适应MQAM技术的频谱利用率和容量之间存在一个固定的功率差，在自适应调制中结合采用网格码调制或点格码调制可以部分地弥补这个功率差。

考虑一族码元间隔固定为T_s的MQAM星座，M代表星座点数。假设采用理想奈奎斯特脉冲，则有$T_s=1/B$，此时平均E_s/N_0等于平均信噪比

$$\frac{\overline{E}_s}{N_0}=\frac{\overline{P}T_s}{N_0}=\overline{\gamma} \tag{9-5}$$

式中各个记号的含义已经在前面的系统模型中给出。给定M时的频带利用率等于每个符号携带的比特数，$R/B=\log_2 M$。自适应系统中的频带利用率一般与平均发送功率$\overline{P}$和误比特率有关。

9.3.1 错误率界

文献[19]给出了AWGN信道中MQAM理想相干解调的误码率界为

$$P_b\leqslant 2\mathrm{e}^{-1.5\gamma/(M-1)}. \tag{9-6}$$

当$M\geqslant 4$且$0\leqslant\gamma\leqslant 30$dB时，一种误差可以小于1dB的误码率紧界为

$$P_b\leqslant 0.2\mathrm{e}^{-1.5\gamma/(M-1)}. \tag{9-7}$$

注意这些只是界，和表6-1给出的错误率公式不完全一样。应用这些界是因为它容易求逆，易于把M写成目标P_b及功率控制的函数。自适应调制中也可以采用不易求逆的BER公式或者仿真得到的BER曲线，再通过数值方法求逆，得到给定BER时的星座大小及信噪比。

289

对于固定功率和速率的非自适应传输，衰落信道下的平均BER是AWGN信道的BER对衰落分布的积分。根据这个表达式，我们可以得出给定平均信噪比和平均BER条件下的最大数据速率，也可以得到给定数据速率和平均BER条件下所需要的平均信噪比，如下面这个例子。

例9.3：考虑瑞利衰落信道及非自适应的BPSK调制，求平均BER为$\overline{P}_b=10^{-3}$所需要的平均信噪比及对应的频带利用率。

解：由6.3.2节可知，瑞利信道下BPSK的$\overline{P}_b\approx 1/4\overline{\gamma}$。因此，对于非自适应传输，达到平均误比特率$\overline{P}_b=10^{-3}$所需要的$\overline{\gamma}=1/4\overline{P}_b=250=24$dB。相应的频带利用率为$R/B=\log_2 2=1$bit/s/Hz。后面将会看到，在相同的信噪比和目标BER条件下，自适应调制的频带利用率要高得多。

9.3.2 自适应速率及功率

考虑根据γ来调整发送功率$P(\gamma)$的情形，约束条件是：平均功率为$\overline{P}$，瞬时BER为$P_b(\gamma)=P_b$。由于接收信噪比成为$\gamma P(\gamma)/\overline{P}$，因此式（9-7）变成

$$P_b(\gamma)\leqslant 0.2\exp\left[\frac{-1.5\gamma}{M-1}\frac{P(\gamma)}{\overline{P}}\right] \tag{9-8}$$

我们要调整M和$P(\gamma)$以满足目标P_b。整理式（9-8）可得到给定P_b时的最大星座为

$$M(\gamma) = 1 + \frac{1.5\gamma}{-\ln(5P_b)} \frac{P(\gamma)}{\overline{P}} = 1 + K\gamma \frac{P(\gamma)}{\overline{P}} \tag{9-9}$$

式中

$$K = \frac{-1.5}{\ln(5P_b)} < 1 \tag{9-10}$$

使频带利用率最大就是要使下式最大：

$$E[\log_2 M(\gamma)] = \int_0^\infty \log_2\left(1 + \frac{K\gamma P(\gamma)}{\overline{P}}\right) p(\gamma)\,\mathrm{d}\gamma \tag{9-11}$$

约束条件为

$$\int_0^\infty P(\gamma)p(\gamma)\,\mathrm{d}\gamma = \overline{P} \tag{9-12}$$

能使式（9-11）最大的功率控制在形式上和能使容量最大的最佳功控式（4-12）相同，为：

$$\frac{P(\gamma)}{\overline{P}} = \begin{cases} 1/\gamma_0 - 1/\gamma K & \gamma \geqslant \gamma_0/K \\ 0 & \gamma < \gamma_0/K \end{cases} \tag{9-13}$$

其中γ_0/K表示最佳截止衰落深度，低于此值时不使用信道。K由式（9-10）给出。定义$\gamma_K = \gamma_0/K$，并在式（9-13）两边同乘以K得到

$$\frac{KP(\gamma)}{\overline{P}} = \begin{cases} 1/\gamma_K - 1/\gamma & \gamma \geqslant \gamma_K \\ 0 & \gamma < \gamma_K \end{cases} \tag{9-14}$$

其中γ_K是截止衰落深度，低于此值时不使用信道。它必须满足功率约束条件（9-12）：

290

$$\int_{\gamma_K}^\infty \left(\frac{1}{\gamma_K} - \frac{1}{\gamma}\right) p(\gamma)\,\mathrm{d}\gamma = K \tag{9-15}$$

将式（9-13）和式（9-14）代入式（9-9），可得瞬时速率为

$$\log_2 M(\gamma) = \log_2\left(\frac{\gamma}{\gamma_K}\right) \tag{9-16}$$

相应的平均频带利用率式(9-11)为：

$$\frac{R}{B} = \int_{\gamma_K}^\infty \log_2\left(\frac{\gamma}{\gamma_K}\right) p(\gamma)\,\mathrm{d}\gamma \tag{9-17}$$

将式（4-12）、式（4-13）同式（9-13）、式（9-17）进行比较可见，能达到容量的最佳自适应系统和最佳的自适应MQAM有相同的功率控制和速率控制，所得到的平均频带利用率也相同，只是自适应MQAM的有效功率损失了K。这个损失量与衰落分布无关。因此，如果衰落信道达到容量Rbit/s/Hz需要的平均信噪比为$\bar{\gamma}$，则要达到相同的速率未编码的自适应MQAM需要的平均接收信噪比为$\bar{\gamma}/K$。也就是说，这种速率和功率都可变的MQAM中，最大可能的编码增益是K。在这种自适应MQAM中叠加网格码或点

格码能获得一部分编码增益，9.3.8节将对此作进一步讨论。

图9-2和图9-3分别画出了对数正态阴影衰落和瑞利衰落下，目标P_b为10^{-3}和10^{-6}时自适应MQAM的平均频带利用率，图中同时画出了容量以便比较。可以注意到自适应MQAM的频谱利用率和容量之间的固定差K，它是误比特率的简单函数。

291

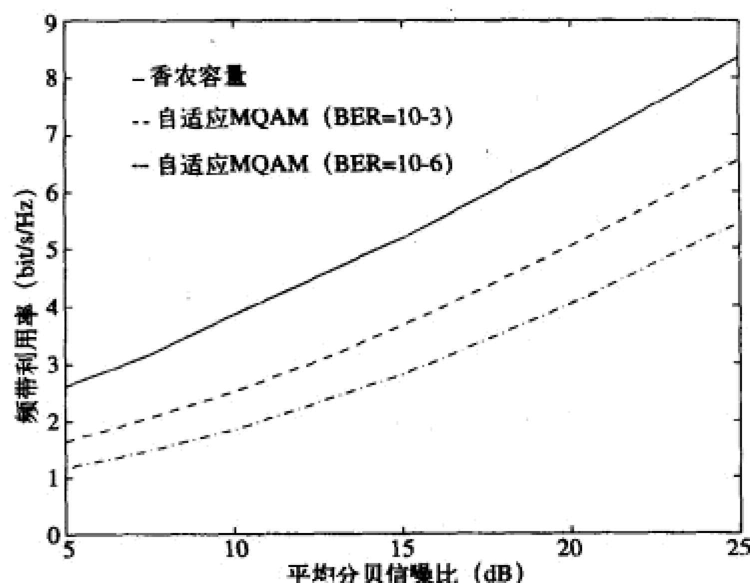


图9-2 对数正态阴影下的平均频带利用率（$\sigma_{\psi_{dB}}=8$dB）

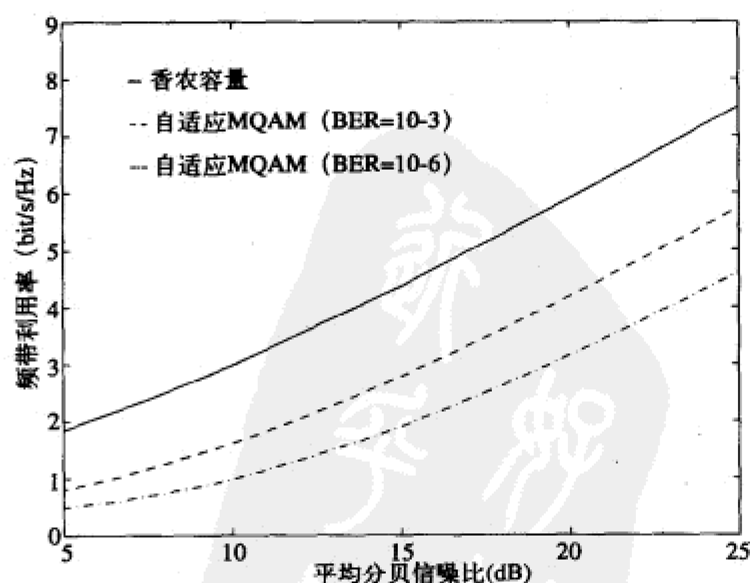


图9-3 瑞利衰落下的平均频带利用率

PDG

9.3.3 固定速率的信道反转

同样可以采用信道反转的功率控制以在接收端维持一个固定的接收信噪比。然后用固定速率的MQAM调制发送就能达到目标P_b。满足式（9-2）的信道反转功控为$P(\gamma)/\overline{P}=\sigma/\gamma$，其中$\sigma=1/E[1/\gamma]$。代入式（9-9）可得给定目标$P_b$时对应的星座大小$M$。由于$R/B=\log_2 M$，因此信道反转功率控制的频谱利用率为

$$\frac{R}{B}=\log_2\left(1+\frac{-1.5}{\ln(5P_b)\,E[1/\gamma]}\right) \tag{9-18}$$

这个结果是基于式（9-7）的，如果得到的$M=2^{R/B}<4$，那么必须用式（9-6），此时需将式（9-18）中的$\ln(5P_b)$替换为$\ln(0.5P_b)$。

使用截断式信道反转时，只有当$\gamma>\gamma_0$时才使用信道，因此将$P(\gamma)/\overline{P}=\sigma/\gamma(\gamma>\gamma_0)$代入式（9-9），并乘以$\gamma>\gamma_0$的概率就可得到截断式信道反转的频谱利用率。优化截止电平γ_0可得到最大的频谱利用率为：

292

$$\frac{R}{B}=\max_{\gamma_0}\log_2\left(1+\frac{-1.5}{\ln(5P_b)\,E_{\gamma_0}[1/\gamma]}\right)p(\gamma>\gamma_0) \tag{9-19}$$

对于目标误比特率为10^{-3}时的瑞利信道，图9-4示出了采用最佳的注水法及截断式信道反转这两种功控时，自适应MQAM的频谱利用率，并画出了两种功控下的容量。出人意料的是，固定速率的截断式信道反转的频谱利用率和速率功率都可变的最佳MQAM大致相同。这说明截断式信道反转在实际中更有用，它不用控制速率就可以达到同速率可变时大致一样的频谱利用率。不过前述的推导对星座大小没有限制，就是说对应式（9-17）、式（9-18）和式（9-19）的M可以为任意实数，并且功率和速率可

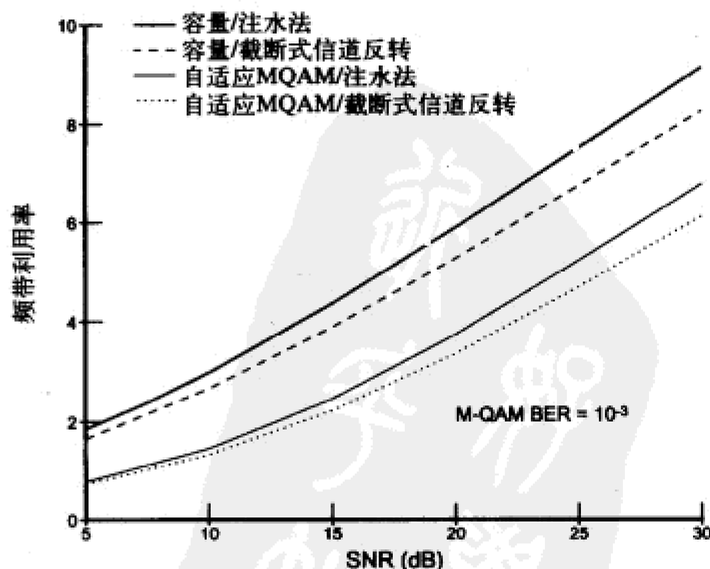


图9-4　不同功控方案下的频谱利用率（瑞利衰落）

以随着γ的变化而连续变化。非整数M的MQAM调制固然有可能，但非常复杂[20]。此外，对发送功率和星座大小实施连续控制也有困难，尤其对于快衰落。因此，我们下面考虑限制星座大小到只是某几个值的情形，这样的限制明显然会影响频谱利用率，但影响不是很大。

9.3.4 离散速率的自适应

假设模型同前，但MQAM的星座图只能在N个正方形星座中选择，其大小依次为$M_0=0$、$M_1=2$及$M_j=2^{2(j-1)}(j=2,\cdots,N-1)$。对$M>2$采用正方形星座是考虑到它更容易实现[21]。我们首先分析这种限制对最佳自适应方案的影响，再分析对信道反转方案的影响。

考虑一个速率和功率都可变的MQAM，在每个码元间隔内，我们用集合$\{M_j:j=0,1,\cdots,N-1\}$中的某一个星座图来发送符号。星座图的选择取决于该码元间隔内的衰落值γ，选择M_0表示不发送。对每个γ值，我们需要决定应该用哪个星座图来传输，以及相应的发送功率是多少。发送端改变发送星座和功率的速度问题将在后面讨论。因为功率连续，星座离散，因此这一方案也称为功率连续速率离散的自适应方案。

293

为了确定每个γ值对应的星座大小，我们将信道的衰落幅度量化。将γ的范围划分成N个衰落区域$R_j=[\gamma_{j-1},\gamma_j)$, $j=0,\cdots,N-1$,其中$\gamma_{-1}=0$，$\gamma_{N-1}=\infty$。当$\gamma\in R_j$时，使用星座图M_j来发送。因此当$j>0$时，$\gamma\in R_j$对应的频谱利用率为$\log_2 M_j$bit/s/Hz。

设计这样的自适应MQAM需要确定出R_j的边界。虽然通过对边界进行优化可以使频谱利用率最大（如9.4.2节），但这样的最佳边界并无闭式解，需要穷举搜索才能得到。故此我们用次最佳的方法来确定边界。这样做要容易得多，而且性能和最佳边界差不多。定义

$$M(\gamma)=\frac{\gamma}{\gamma_K^*} \tag{9-20}$$

式中$\gamma_K^*>0$是一个参数，后面我们会优化这个参数以最大化频谱利用率。注意将式（9-13）代入式（9-9）就是$\gamma_K^*=\gamma_K$时的式（9-20），因此，选择适当的γ_K^*值，式（9-20）就表示无星座限制时的最佳星座大小。

先假设γ_K^*是固定值，并定义$M_N=\infty$。为了确定出给定信噪比为γ时应该选择$M_j(j=0,1,\cdots,N-1)$中的哪个星座，我们根据式（9-20）计算出$M(\gamma)$，再找出这样一个j使得$M_j\leqslant M(\gamma)<M_{j+1}$，将这个$\gamma$值对应到星座$M_j$。就是说，对于给定的$\gamma$，我们用$\{M_j:j=0,\cdots,N-1\}$中小于$M(\gamma)$的最大星座进行发送。例如衰落幅度$\gamma$满足$2\leqslant\gamma/\gamma_K^*<4$时，用BPSK。这样，除去$\gamma_{-1}=0$和$\gamma_{N-1}=\infty$之外的其他边界位于$\gamma_j=\gamma_K^*M_{j+1}$，$j=0,\cdots,N-2$。很明显，增加星座图的数量$N$可以更好地近似于式（9-9）中的连续自适应方案，从而能得到更高的频谱利用率。

各个区域和它们所对应的星座确定了以后，还需要确定出能满足误比特率要求和功率约束要求的功率控制方案。对于星座$M_j>0$要维持固定的误比特率，由式（9-9）可知这样的功控策略是：若$\gamma\in R_j$，则

$$\frac{P_j(\gamma)}{\bar{P}}=\begin{cases}(M_j-1)(1/\gamma K) & M_j<\gamma/\gamma_K^*\leqslant M_{j+1}\\ 0 & M_j=0\end{cases} \tag{9-21}$$

因为它在星座M_j下的E_s/N_0为：

$$\frac{E_s(j)}{N_0}=\frac{\gamma P_j(\gamma)}{\bar{P}}=\frac{M_j-1}{K} \tag{9-22}$$

根据K的定义可知，星座大小为M_j、E_s/N_0为式（9-22）的MQAM的误比特率就是所期望的目标值P_b。表9-1列出了五个衰落区域时，星座大小和功率控制与γ和γ_K^*的关系。

294

表9-1　五区域的速率和功率自适应

区域（j）	γ范围	M_j	$P_j(\gamma)/\bar{P}$
0	$0\leqslant\gamma/\gamma_K^*<2$	0	0
1	$2\leqslant\gamma/\gamma_K^*<4$	2	$1/K\gamma$
2	$4\leqslant\gamma/\gamma_K^*<16$	4	$3/K\gamma$
3	$16\leqslant\gamma/\gamma_K^*<64$	16	$15/K\gamma$
4	$64\leqslant\gamma/\gamma_K^*<\infty$	64	$63/K\gamma$

这种离散速率方案的频谱利用率是各区域的速率按其出现概率进行加权后的和：

$$\frac{R}{B}=\sum_{j=1}^{N-1}\log_2(M_j)p\left(M_j\leqslant\frac{\gamma}{\gamma_K^*}<M_{j+1}\right) \tag{9-23}$$

因为M_j是γ_K^*的函数，故可按γ_K^*最大化式（9-23），约束条件为

$$\sum_{j=1}^{N-1}\int_{\gamma_K^*M_j}^{\gamma_K^*M_{j+1}}\frac{P_j(\gamma)}{\bar{P}}p(\gamma)\,\mathrm{d}\gamma=1 \tag{9-24}$$

其中$P_j(\gamma)/\bar{P}$由式(9-21)给出。最佳的γ_K^*没有闭式解，下面计算中的最佳值是用数值搜索得到的。

图9-5和图9-6分别示出了对数正态阴影衰落和瑞利衰落下，衰落区域数N与式（9-23）的最大值之间的关系，图中误比特率均为10^{-3}。从图9-5中可以看出，在对数正态阴影衰落下，六个衰落区域（$M_j=0, 2, 4, 16, 64, 256$）的频谱利用率与连续速率相比，相差不到1dB。在瑞利衰落下，五区域（$M_j=0, 2, 4, 16, 64$）的结果也可以得到类似结论。

上述的离散速率自适应可通过固定每个星座图M_j对应的发送功率进一步简化，即每一个衰落区域对应一个星座图和一个发送功率。这就是离散功率离散速率的自适应MQAM。对应每个区域，信道信噪比γ有一定的变化范围，但发送功率和星座大小是固定的，所以误比特率是随γ而变化的。因此，在确定区域边界和发送功率时，应采用平均BER目标。

295

对可用星座图的上述制约对完全信道反转和截断式信道反转也有影响。假设星座图只能从集合$\mathcal{M}=\{M_0=0,\cdots,M_{N-1}\}$中选择，则完全信道反转的频谱利用率为

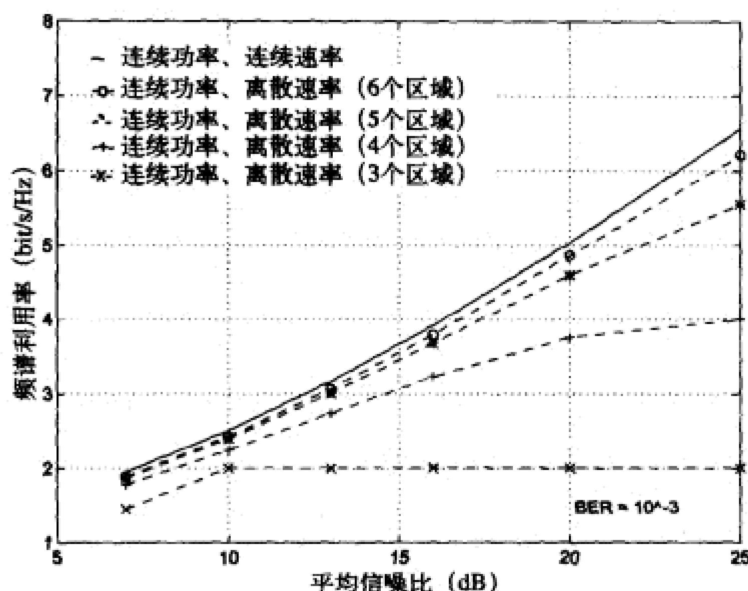


图9-5 对数正态阴影下离散速率控制的频谱利用率（$\sigma_{\psi_{dB}}=8$dB）

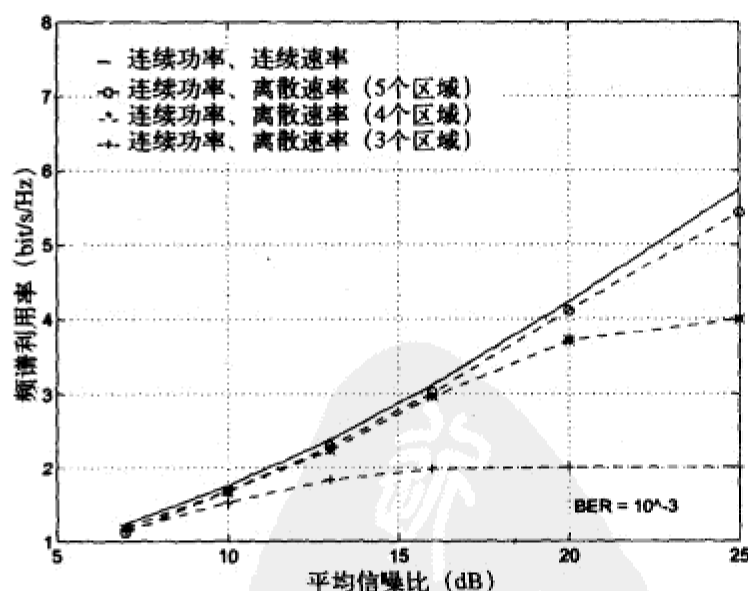


图9-6 瑞利衰落下离散速率控制的频谱利用率

$$\frac{R}{B}=\log_2\left\lfloor\left(1+\frac{-1.5}{\ln(5P_b)E[1/\gamma]}\right)\right\rfloor_{\mathcal{M}} \tag{9-25}$$

式中$\lfloor x\rfloor_{\mathcal{M}}$表示集合$\mathcal{M}$中小于或等于$x$的最大值。这种方案的频谱效率只能取值于$\log_2 M(M\in\mathcal{M})$，它是一个阶跃函数，跳变点处的信噪比$\bar{\gamma}$正好使无星座限制的频谱效率式(9-18)等于$\log_2 M$。截断式信道反转方案的频谱利用率为

$$\frac{R}{B} = \max_{\gamma_0} \log_2 \left[\left(1 + \frac{-1.5}{\ln(5P_b)\, E_{\gamma_0}[1/\gamma]} \right) \right]_{\mathcal{M}} p(\gamma > \gamma_0) \qquad (9\text{-}26)$$

图9-7和图9-8示出了星座限制对不同功率控制的影响。图中的星座限制是$\mathcal{M}=\{M_0=0, 2, 4, \cdots, 256\}$，有六个衰落区域。离散功率离散速率时的平均BER等于离散速率连续功率时的瞬时BER。从图中可见，对于速率离散可变的MQAM，进一步使每种星座下的发送功率固定为一个值将造成大约1~2dB的性能损失。图中还画出了式（9-17）给

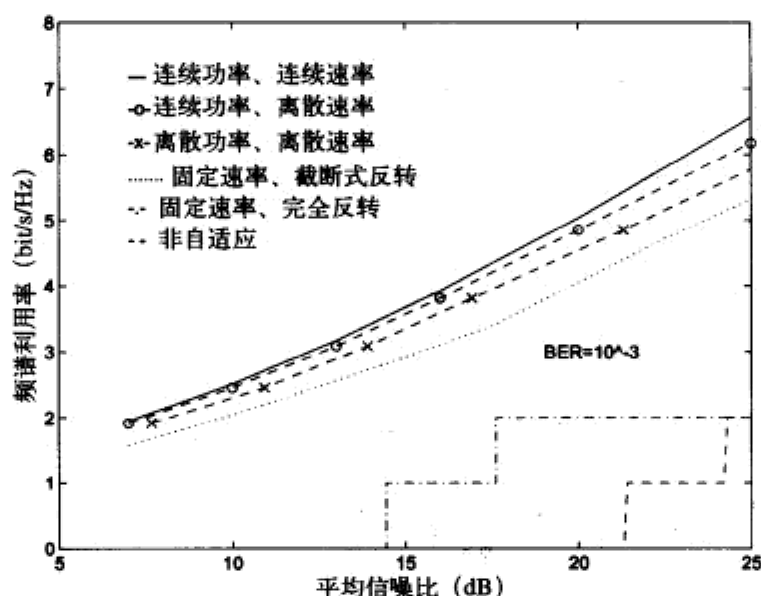


图9-7　对数正态阴影下的频谱利用率（$\sigma_{\psi_{dB}}=8\text{dB}$）

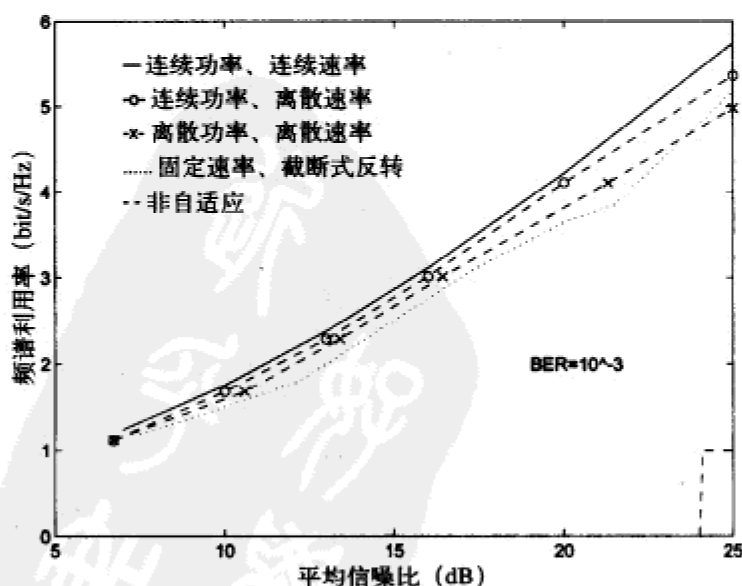


图9-8　瑞利衰落下的频谱利用率

出的连续功率、连续速率自适应的最大频谱利用率，所有离散速率的方案与这个理论上最佳的曲线相比，性能损失在3dB以内。 296

图中还示出了式（9-26），即固定速率的截断式信道反转的频谱利用率。它和离散功率的频谱利用率非常接近，但这个频谱效率对应的γ_0非常大，相应的中断率$P_{out}=p(\gamma\leqslant\gamma_0)$在0.1~0.6范围内变化。它在信道条件较好时以高数据速率进行突发传输，有点像分组无线电。图9-7中还画出了对数正态阴影下完全信道反转的频谱利用率，即式（9-25）。对于瑞利衰落，完全反转的频带利用率是零。图9-7和图9-8中还画出了传输速率和传输功率都是固定的非自适应传输的频谱利用率，此频谱利用率是这样得到的：对式（9-31）中的错误率按衰落分布$p(\gamma)$进行平均可得到平均错误率，再根据$\bar{\gamma}$和目标平均误比特率10^{-3}确定出M的值，如例9.3。非自适应传输虽然非常简单，但其频谱效率非常差。在信道变化很快、精确估计信道很难时，非自适应的传输也许是最好的选择。将目标错误率设为10^{-6}也可以得类似的曲线，此时的频带利用率和目标错误率为10^{-3}时相比存在一定的损失，损失量大致和图9-2、图9-3中所示的损失量差不多。 297

9.3.5 衰落区的平均持续时间

自适应调制中，衰落区个数的选择取决于信道的变化速度和硬件的限制。它决定了发送端有多少星座可以选择，决定了发送端改变星座和功率的频度。如果考虑信道估计、反馈以及硬件限制等因素的话，一般要求星座图在数十或数百码元周期内保持不变。另外，线性功放和带外辐射等因素也不允许发送功率变化过快。有关硬件实现的深入讨论参见文献[22]。我们现在来考虑信噪比γ滞留在某一特定衰落区域R_j中的持续时间，它涉及如何在区域数和发送功率/星座的变化速度之间进行取舍的问题。现在我们来讨论给定衰落区域内信噪比的持续时间。

令$\bar{\tau}_j$表示γ滞留在第j个衰落区的平均持续时间，令$A_j=\gamma_K^*M_j$，其中γ_K^*和M_j的含义同前。第j个衰落区为$\{\gamma:A_j\leqslant\gamma<A_{j+1}\}$。称$\bar{\tau}_j$为第$j$个平均衰落区时间（average fade region duration，AFRD）。这个概念类似于3.2.3节中的平均衰落时间（average fade duration，AFD）。不同之处是，AFD是γ滞留在某个电平之下的平均时间，而AFRD是γ在滞留在两个电平之间的平均时间。对最差的那个衰落区（$j=0$），这两个概念相同。

$\bar{\tau}_j$的准确值需要通过联合密度$p(\gamma,\dot{\gamma})$进行复杂的推导，这还是一个尚未解决的问题。不过用3.2.4节中的有限状态马尔可夫模型可以使问题得到很好的近似。该模型按时间间隔T（通常为码元周期）将衰落近似为一个离散时间的马尔可夫过程。它假设在每个码元周期内衰落值γ保持在同一衰落区内，下一码元周期内，只能转移到本衰落区或邻近的衰落区。注意，这个近似模型得到的深衰落持续时间要比精确模型的结果长[23]。在上述假设下，衰落区之间的转移概率为：

$$p_{j,j+1}=\frac{L_{j+1}T}{\pi_j},\quad p_{j,j-1}=\frac{L_jT}{\pi_j},\quad p_{j,j}=1-p_{j,j+1}-p_{j,j-1} \tag{9-27}$$

298

其中L_j是A_j的电平通过率，π_j是第j个衰落区的稳态分布：$\pi_j=p(A_j\leqslant\gamma<A_{j+1})$。由于马尔可夫过程停留在给定的状态的时间服从几何分布（见文献[24]的2.3节），$\bar{\tau}_j$为：

$$\bar{\tau}_j = \frac{T}{p_{j,j+1} + p_{j,j-1}} = \frac{\pi_j}{L_{j+1} + L_j} \tag{9-28}$$

$\bar{\tau}_j$是电平通过率和衰落分布的简单函数。瑞利衰落的电平通过率是已知的（见文献[25]的1.3.4节），对数正态阴影衰落的电平通过率未知，因为其联合分布 $p(\gamma,\dot{\gamma})$未知。

式（3-44）给出的瑞利衰落的电平通过率为：

$$L_j = \sqrt{\frac{2\pi A_j}{\bar{\gamma}}} f_D e^{-A_j/\bar{\gamma}} \tag{9-29}$$

式中$f_D = v/\lambda$是多普勒频率。将式（9-29）代入式（9-28）可知$\bar{\tau}_j$跟多普勒频率成反比。由于π_j和A_j与f_D无关，所以，如果已知多普勒频率为f_D时的$\bar{\tau}_j$，那么多普勒频率为$\hat{f}_D$时的$\bar{\hat{\tau}}_j$为

$$\bar{\hat{\tau}}_j = \frac{f_D}{\hat{f}_D} \bar{\tau}_j \tag{9-30}$$

表9-2列出了$f_D = 100$Hz的瑞利衰落[1]在五个衰落区($M_j = 0, 2, 4, 16, 64$)中的$\bar{\tau}_j$值，平均功率分别为 $\bar{\gamma} = 10\text{dB}(\gamma_K^* = 1.22)$ 和 $\bar{\gamma} = 20\text{dB}(\gamma_K^* = 1.685)$ 。其他多普勒频率的AFRD可利用表中的值和式（9-30）得出。从这个表可以看出，即使对于高速移动的环境，数据速率为100k波特时，离散速率离散功率自适应的星座和发送功率也能在数十到数百码元间隔内保持不变。

表9-2　$f_D = 100$Hz时的平均衰落区持续时间 $\bar{\tau}_j$

区域（j）	$\bar{\gamma}$= 10dB	$\bar{\gamma}$= 20dB
0	2.23ms	0.737ms
1	0.83ms	0.301ms
2	3.00ms	1.06ms
3	2.83ms	2.28ms
4	1.43ms	3.84ms

例9.4：假设瑞利衰落信道的 $\bar{\gamma} = 10$dB，$f_D = 50$Hz，五个衰落区为$M_j = 0, 2, 4, 16, 64$，求AFRD。

解：除了多普勒频率 $\hat{f}_D = 50$Hz 是$f_D = 100$Hz的一半外，本例中的其他参数都跟表9-2一样。因此，由式（9-30），将表中的每个值乘以 $f_D / \hat{f}_D = 2$，即为所求的AFRD。

299

阴影衰落下，可根据式（2-50）给出的自相关函数 $A(\delta) = \sigma_{\psi_{dB}}^2 e^{-\delta/X_c}$ 来粗略估计 $\bar{\tau}_j$，式中的$\delta = v\tau$，v是移动台的速度。比如可以近似成 $\bar{\tau}_j \approx 0.1X_c / v$，这时，间隔为$\bar{\tau}_j$的相关系数是0.9，如果区域数不是很多的话，在$\bar{\tau}_j$时间内$\gamma$的变化极有可能还在同一衰落区内。

9.3.6　误码率近似的影响

前几节介绍的自适应策略是在9.3.1节给出的BER界的基础上导出的。因为是上界，

1. 文献[26]研究了有限状态马尔可夫模型对瑞利信道的有效性问题。

所以，实际BER可能比目标值低。下面我们来看实际BER和目标值到底差多少。系统实际达到的BER可以通过仿真或更好的近似式来得到。根据表6-1，戈雷码映射的MQAM在高信噪比下的BER可以很好地近似为

$$P_b \approx \frac{4}{\log_2 M} Q\left(\sqrt{\frac{3\gamma}{M-1}}\right) \tag{9-31}$$

对于连续功率离散速率的自适应调制，第j个星座的符号信噪比γ为

$$\frac{E_s(j)}{N_0} = \frac{M_j - 1}{K} \tag{9-32}$$

将式（9-31）按每个星座图进行平均可得到一个比较准确的平均BER解析式，为：

$$\bar{P}_b = \frac{\sum_{j=1}^{N-1} 4Q\left(\sqrt{3/K}\right) \int_{\gamma_K^* M_j}^{\gamma_K^* M_{j+1}} p(\gamma)\,\mathrm{d}\gamma}{\sum_{j=1}^{N-1} \log_2 M_j \int_{\gamma_K^* M_j}^{\gamma_K^* M_{j+1}} p(\gamma)\,\mathrm{d}\gamma} \tag{9-33}$$

其中$M_N = \infty$。

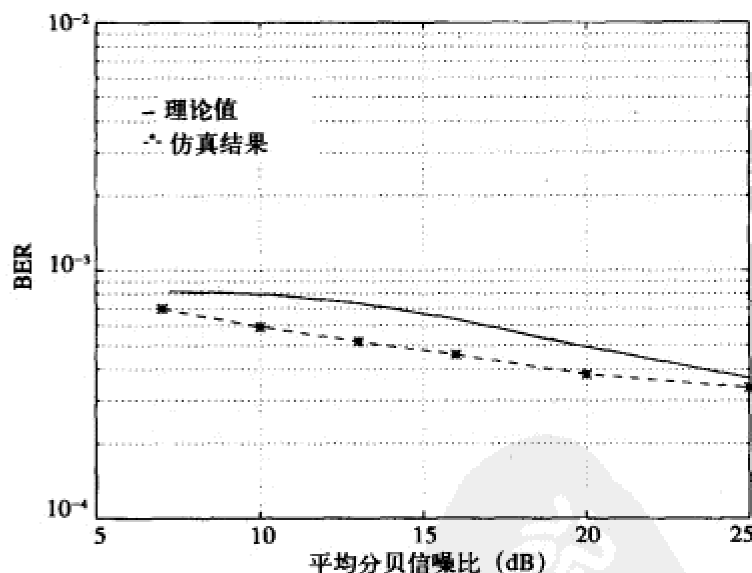


图9-9 对数正态阴影下的误比特率（六个衰落区）

对于功率可变速率可变的自适应MQAM以及目标BER为10^{-3}，图9-9和图9-10分别对于对数正态阴影衰落和瑞利衰落两种情形，画出了由解析式（9-33）得到的BER和实际仿真得到的BER。仿真得到的BER比式（9-33）的理论值稍好一些，原因在于式(9-33)是在最近邻界中取了最大的近邻数。当$\bar{\gamma} > 10$dB时，仿真结果和理论结果都小于目标误比特率10^{-3}。在低信噪比时，误比特率超过了目标值10^{-3}，这是因为信噪比低的时候，自适应调制使用BPSK的机会增大，但式（9-7）不适用于BPSK，使得此时的P_b大于式（9-7）所认为的值，这种情况下应当使用较松的界（9-6）。高信噪比时仿真得到的BER低于目

标值，说明图9-5和图9-6给出的理论结果偏于保守，我们有可能在保持目标P_b仍为10^{-3}的

300

条件下，还能稍微提高频谱利用率。

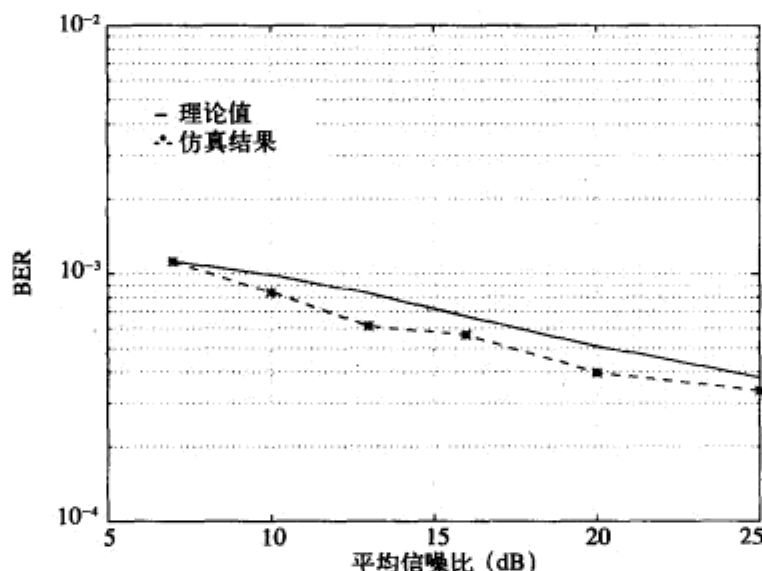


图9-10 瑞利衰落下的误比特率（五个衰落区）

9.3.7 信道估计误差和估计延时的影响

本节考虑估计误差和延时带来的影响，此时估计误差$\varepsilon=\hat{\gamma}/\gamma\neq 1$，延时$i_d=i_f+i_e\neq 0$。首先考虑估计误差。假设发送端不是用真实值$\gamma$，而是用其估计值$\hat{\gamma}$来控制发送功率和速率。由式（9-8），误比特率的界为

$$P_b(\gamma,\hat{\gamma})\leqslant 0.2\exp\left[\frac{-1.5\gamma}{M(\hat{\gamma})-1}\frac{P(\hat{\gamma})}{\bar{P}}\right]=0.2[5P_b]^{1/\varepsilon} \tag{9-34}$$

最右边的等式是通过带入最佳速率式（9-9）和最佳功率式（9-13）得到的。当$\varepsilon=1$时，式（9-34）退化为目标P_b。当$\varepsilon\neq 1$时，$\varepsilon>1$使BER增大，高于目标值，$\varepsilon<1$使BER减小。

估计误差使误比特率成为

301

$$\bar{P}_b\leqslant\int_0^{\infty}\int_{\gamma_K}^{\infty}0.2[5P_b]^{\gamma/\hat{\gamma}}p(\gamma,\hat{\gamma})\,\mathrm{d}\hat{\gamma}\,\mathrm{d}\gamma=\int_0^{\infty}0.2[5P_b]^{1/\varepsilon}p(\varepsilon)\,\mathrm{d}\varepsilon \tag{9-35}$$

概率密度函数$p(\varepsilon)$与联合概率密度函数$p(\gamma,\hat{\gamma})$有关，后者又取决于具体的信道估计技术。文献[27]指出，采用导频估计时，信号包络和估计值的联合分布是双变量瑞利分布。文献[27]用这个联合分布来求解有信道估计误差时非自适应调制系统的错误率。其分析方法可以扩展到自适应调制中。

如果估计误差不超过一个有限范围内的话，我们可以用式（9-34）得到其影响的上界。图9-11示出误差ε为常量时BER的增量。从这个图看，目标BER为10^{-3}时，估计误差应当小于1dB；目标BER为10^{-6}时，估计误差应当小于0.5dB。因为假设信道误差为常

数，所以这个判断是偏于保守的。即便如此，利用文献[28]中提出的导频估计技术，通过选择适当的参数，可以使估计误差保持在一定的范围内。当信道被低估时($\varepsilon<1$)，BER会减小，但频谱利用率将有一些损失，这是因为估计值的均值$\bar{\hat{\gamma}}$和真实值的均值$\bar{\gamma}$不一样。文献[29]中讲到了平均功率估计误差的影响。

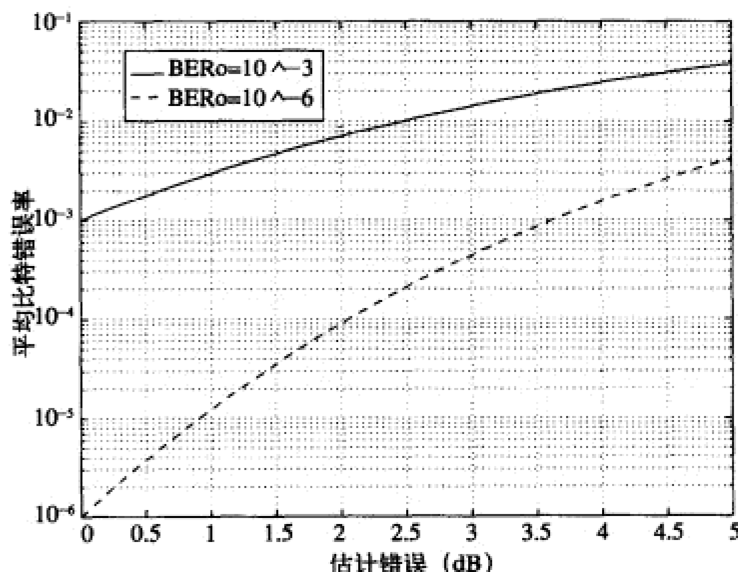


图9-11 估计错误给BER带来的影响

现在假设信道估计完全正确（$\varepsilon=1$），但估计和反馈引起的时延i_d不为零。在时刻i，发送端使用滞后的信道估计$\hat{\gamma}[i]=\gamma[i-i_d]$来调整发送功率和发送速率。文献[30]中指出，以延时后的信道值为条件时，接收信号服从莱斯分布，对这个分布进行平均可得到错误率。文献[30]还进一步提出了一种自适应编码方法，可以减轻估计时延对自适应调制的影响。除此之外，采用信道预测也可以减轻这种影响[31]。

302

也可以用和式（9-34）类似的方法得到估计时延引起的BER增量。给定信道信噪比的真实值为$\gamma[i]$，延迟值为$\gamma[i-i_d]$，我们有

$$\begin{aligned}P_b(\gamma[i],\gamma[i-i_d]) &\leqslant 0.2\exp\left[\frac{-1.5\gamma[i]}{M(\gamma[i-i_d])-1}\frac{P(\gamma[i-i_d])}{\bar{P}}\right] \\ &=0.2[5P_{b_0}]^{\gamma[i]/\gamma[i-i_d]}\end{aligned} \tag{9-36}$$

定义$\xi[i,i_d]=\gamma[i]/\gamma[i-i_d]$。因为$\gamma[i]$是平稳遍历的，所以给定$\gamma[i]$时，$\xi[i,i_d]$的分布只和$i_d$及$\gamma=\gamma[i]$的值有关，记其为$p_{i_d}(\xi|\gamma)$。对$\xi$和$\gamma$积分便可得到平均BER。文献[32]具体给出BER为

$$P_b[i_d]=\int_{\gamma_K}^{\infty}\left[\int_0^{\infty}0.2[5P_{b_0}]^{\xi}p_{i_d}(\xi\mid\gamma)\,\mathrm{d}\xi\right]p(\gamma)\,\mathrm{d}\gamma \tag{9-37}$$

式中γ_K是最佳功控的截止门限，$p(\gamma)$是衰落分布。$p_{i_d}(\xi|\gamma)$取决于衰落过程的自相关函数。文献[32]给出了Nakagami衰落（瑞利衰落是其特例）下，$p_{i_d}(\xi|\gamma)$的闭式解。将这个分布

代入式（9-37）可得到瑞利衰落下的平均BER，它是时延参数i_d函数，图9-12按归一化时延$i_d f_d$画出了这个结果。从图中可知，为了使BER接近目标值，由信道估计过程和反馈过程造成的总时延应限制在$0.001/f_D$内。

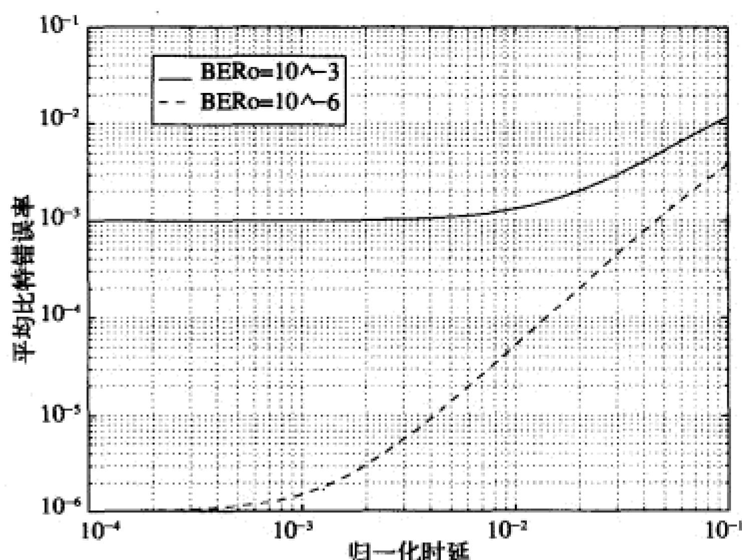


图9-12 归一化时延（$i_d f_d$）对误比特率的影响

9.3.8 自适应编码调制

在自适应调制中叠加上网格码或点格码能获得额外的编码增益。通过编码调制固有的子集划分，为AWGN信道设计的网格码（或点格码）可以直接叠加到自适应调制中，并具有相同的编码增益。如8.7节所述，编码调制中编码设计和星座设计是单独进行的，303 自适应编码调制正是利用了这一点。

由于编码调制中的信道编码增益基本上独立于调制，所以编码调制对于速率和功率都可变的MQAM来说是一种自然的选择。我们可以根据瞬时信噪比调整发送功率和调制电平数（星座点数），而不影响信道编码增益。

在图9-13所示的编码调制框图中，陪集码的设计和AWGN信道时相同，即点格结构和卷积编码器同8.7节中的网格或点格编码调制。调制器中的星座图采用正方形点格，其星座点数M可调。MQAM星座图的大小由发送功率决定，发送功率的决定方式如同9.3.2节中无编码的情形，取决于瞬时信噪比和所需的误比特率。

用G_c表示式（8-87）给出的编码调制的编码增益。对式（9-7）的BER近似式根据编码增益进行调整，则对于给定的信噪比γ，有

$$P_b \approx 0.2\mathrm{e}^{-1.5(\gamma G_c/M-1)} \tag{9-38}$$

式中M是发送星座图的大小。如同未编码时的情形，基于式（9-7）这个误码率紧界，我们可根据瞬时信噪比来调整星座图大小M以及发送的信号功率，以维持给定的误比

特率：

$$M(\gamma)=1+\frac{1.5\gamma G_c}{-\ln(5P_b)}\frac{P(\gamma)}{\bar{P}} \tag{9-39}$$

用于选择陪集点的未编码的比特数为：

$$n(\gamma)-2k/N=\log_2 M(\gamma)-2(k+r)/N.$$

这个值是随时间变化的，因此需要对未编码比特进行缓冲排队，如图9-13所示。 304

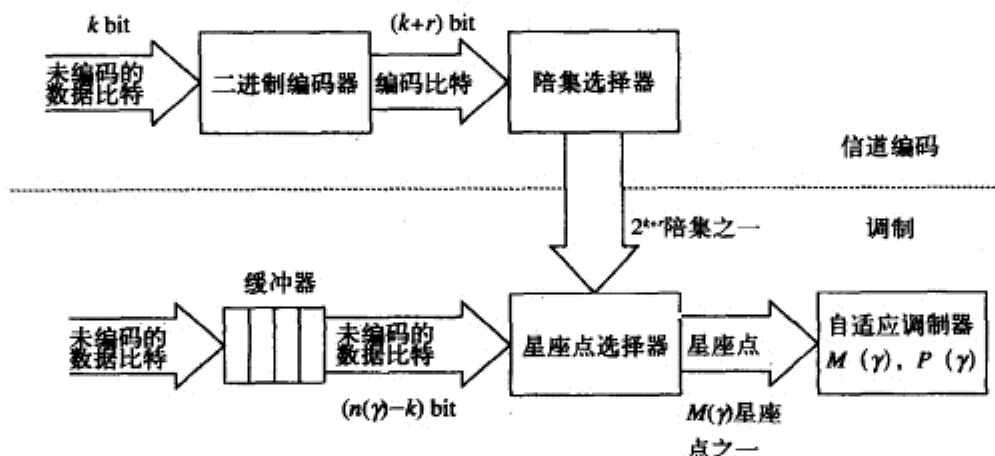


图9-13 自适应编码调制

每次发送的比特速率为$\log_2 M(\gamma)$，数据速率为$\log_2 M(\gamma)-2r/N$。可在给定的平均功率约束下，通过最大化$E=[\log_2 M]$来最大化数据速率。这样得到的最佳功控方案为：

$$\frac{P(\gamma)}{\bar{P}}=\begin{cases}1/\gamma_0-1/\gamma K_c & \gamma\geqslant\gamma_0/K_c\\ 0 & \gamma<\gamma_0/K_c\end{cases} \tag{9-40}$$

式中$\gamma_{K_c}=\gamma_0/K_c$是截止衰落深度，$K_c=KG_c$，$K$由式（9-10）给出。这个结果实际就是用$K_c$代替无编码的功控式（9-13）中的$K$。因此，自适应的编码调制相对于未编码的自适应调制有G_c的增益。将式（9-40）代入式（9-39）可得速率控制为

$$M(\gamma)=\left(\frac{\gamma}{\gamma_{K_c}}\right) \tag{9-41}$$

总的频谱利用率为

$$\frac{R}{B}=\int_{\gamma_{K_c}}^{\infty}\log_2\left(\frac{\gamma}{\gamma_{K_c}}\right)p(\gamma)\,\mathrm{d}\gamma \tag{9-42}$$

如果G_c未包括星座扩展因子，那么必须从式（9-42）中减去$2r/N$，才能得到数据速率。文献[33]对自适应编码调制有更多的讨论，并给出了各种不同复杂度的自适应网格调制的频谱利用率曲线。其结果表明，在合理的复杂度下，自适应网格编码调制的性能和香农界相差不到5dB，在未编码的自适应调制上叠加网格码得到的编码增益和AWGN信道下网格码的编码增益大致相同。

9.4 一般的M进制调制

前面提到的自适应MQAM也可以推广到其他的M进制调制中。不管调制类型是什么，其基本原理是相同的，就是调整发送功率和星座大小，在保证瞬时BER的前提下，能使平均数据速率最大化。本节讨论采用一般的M进制调制的情况下，连续速率和离散速率的最佳速率控制和最佳功率控制。

9.4.1 连续速率的自适应

首先考虑速率和功率都可以连续调整的情形。我们要做的是，对于一般的M进制调制、给定平均功率$\overline{P}$和目标误比特率，找出能使平均数据速率$E[k(\gamma)]$取得最大值的最佳的功率控制$P(\gamma)$和速率控制$k(\gamma)=\log_2 M(\gamma)$，当BER的精确值或近似值可以写成

305

$$P_b(\gamma)\approx c_1\exp\left[\frac{-c_2\gamma(P(\gamma)/\overline{P})}{2^{c_3k(\gamma)}-c_4}\right] \tag{9-43}$$

的形式时，上述的最优化问题可以得到简化，式中的c_1、c_2和c_3是正常数，c_4是实常数。例如对于式（9-6）和式（9-7）给出的MQAM的BER界，$c_1=2$或0.2、$c_2=1.5$、$c_3=1$、$c_4=1$。通过适当的曲线拟合，多数M进制调制的BER可以近似成这种形式。

式（9-43）的好处是容易求逆，这使我们可以把速率$k(\gamma)$表示为功率控制$P(\gamma)$和目标误比特率P_b的函数：

$$k(\gamma)=\log_2 M(\gamma)=\begin{cases}\dfrac{1}{c_3}\log_2\left[c_4-\dfrac{c_2\gamma}{\ln(P_b/c_1)}\dfrac{P(\gamma)}{\overline{P}}\right] & P(\gamma)\geqslant 0,\ k(\gamma)\geqslant 0\\ 0 & 其他\end{cases} \tag{9-44}$$

为了得到能最大化频谱利用率$E[k(\gamma)]$的功率控制和速率控制，可以建立如下拉格朗日函数：

$$J(P(\gamma))=\int_0^\infty k(\gamma)p(\gamma)\,\mathrm{d}\gamma+\lambda\left[\int_0^\infty P(\gamma)p(\gamma)\,\mathrm{d}\gamma-\overline{P}\right] \tag{9-45}$$

最佳自适应方案必能以非负的速率和功率最大化这个拉格朗日函数，所以它满足

$$\frac{\partial J}{\partial P(\gamma)}=0,\quad P(\gamma)\geqslant 0,\ k(\gamma)\geqslant 0 \tag{9-46}$$

代入式（9-44）中的$k(\gamma)$，可解得最佳功控为

$$\frac{P(\gamma)}{\overline{P}}=\begin{cases}-1/c_3(\ln 2)\lambda\overline{P}-1/\gamma K & P(\gamma)\geqslant 0,\ k(\gamma)\geqslant 0\\ 0 & 其他\end{cases} \tag{9-47}$$

其中

$$K=-\frac{c_2}{c_4\ln(P_b/c_1)} \tag{9-48}$$

式（9-47）中的功率控制可写成更简略的形式：

PDG

$$\frac{P(\gamma)}{\overline{P}}=\begin{cases}\mu-1/\gamma K & P(\gamma)\geqslant 0,\ k(\gamma)\geqslant 0\\ 0 & \text{其他}\end{cases} \tag{9-49}$$

其中的常数μ由平均功率约束条件（9-12）决定。

虽然式（9-49）中的最佳功控表达式看起来很简单，但它的实际表现却与P_b近似式（9-43）中的c_4值有很大的关系。将MQAM的BER近似式(9-6)或(9-7)应用到式（9-43），得到的最佳功控就是式（9-13）给出的注水法。但我们下面将会看到，注水法并不是对所有情形都是最佳的。

根据式（6-16），格雷编码MPSK的BER可以较好地近似为

$$P_b\approx\frac{2}{\log_2 M}Q\left(\sqrt{2\gamma}\sin\left(\frac{\pi}{M}\right)\right) \tag{9-50}$$

不过这个表达式不具有式（9-43）的形式。因为Q函数不易求逆，所以无法直接求解出最佳的速率控制和功率控制。为此，我们考虑MPSK误比特率的三种界，它们适用于$k(\gamma)\geqslant 2$的情形：

$$\text{界1：}\quad P_b(\gamma)\approx 0.05\exp\left[\frac{-6\gamma(P(\gamma)/\overline{P})}{2^{1.9k(\gamma)}-1}\right] \tag{9-51}$$

$$\text{界2：}\quad P_b(\gamma)\approx 0.2\exp\left[\frac{-7\gamma(P(\gamma)/\overline{P})}{2^{1.9k(\gamma)}+1}\right] \tag{9-52}$$

$$\text{界3：}\quad P_b(\gamma)\approx 0.25\exp\left[\frac{-8\gamma(P(\gamma)/\overline{P})}{2^{1.94k(\gamma)}}\right] \tag{9-53}$$

306

图9-14画出了这些界及较精确的近似式（9-50）。可以看出，这些界能较好地近似

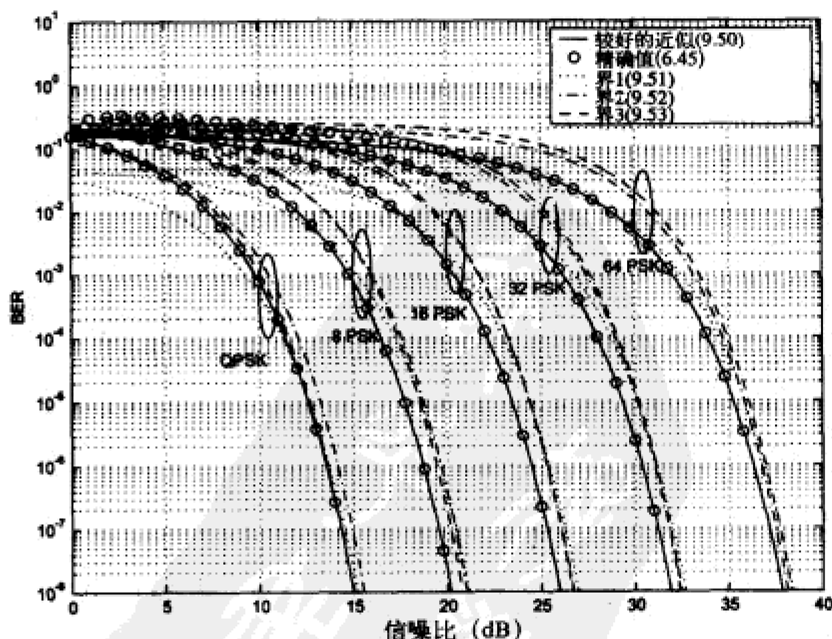


图9-14 MPSK的误比特率界

式（6-45）给出的精确值，特别是在高信噪比时。

对于第一个界（9-51），$c_1=0.05$、$c_2=6$、$c_3=1.9$、$c_4=1$。假设目标P_b小于0.05，则式（9-49）中的$K=-c_2/(c_4\ln(P_b/c_1))$是正值。因此，为了使功率$P(\gamma)/\bar{P}=\mu-1/\gamma K$为正值，$\mu$必为正。此外，$K$为正时，对任何$P(\gamma)\geqslant 0$有$k(\gamma)\geqslant 0$。这样式（9-49）变为：

$$\frac{P(\gamma)}{\bar{P}}=\begin{cases}1/\gamma_0 K-1/\gamma K & P(\gamma)\geqslant 0\\ 0 & \text{其他}\end{cases} \tag{9-54}$$

式中$\gamma_0\geqslant 0$是截止衰落深度，低于此值时不发送信号。这个截止值也和μ一样由平均功率约束条件（9-12）确定。式（9-54）给出的功控和自适应MQAM中的注水法功控式（9-13）相同，这是因为MPSK的BER界（9-51）和MQAM的BER界（9-6）、（9-7）相似。将式（9-54）代入式（9-44）可得到相应的最佳速率控制为：

$$k(\gamma)=\begin{cases}(1/c_3)\log_2(\gamma/\gamma_0) & \gamma\geqslant\gamma_0\\ 0 & \text{其他}\end{cases} \tag{9-55}$$

307 它也和自适应MQAM的速率控制式（9-16）有相同的形式。

现在考虑第二个界（9-52），此时$c_1=0.2$、$c_2=7$、$c_3=1.9$、$c_4=-1$。假设目标$P_b<0.2$，则$K=-c_2/(c_4\ln(P_b/c_1))$是负值。由式（9-44）可知，当K是负值时，必须$\mu\geqslant 0$才能使$k(\gamma)\geqslant 0$，这样，$P(\gamma)\geqslant 0$、$k(\gamma)\geqslant 0$时的最佳功率控制为：

$$\frac{P(\gamma)}{\bar{P}}=\begin{cases}\mu-1/\gamma K & k(\gamma)\geqslant 0\\ 0 & \text{其他}\end{cases} \tag{9-56}$$

再由式（9-44）可得最佳速率控制为

$$k(\gamma)=\begin{cases}(1/c_3)\log_2(\gamma/\gamma_0) & \gamma\geqslant\gamma_0\\ 0 & \text{其他}\end{cases} \tag{9-57}$$

式中$\gamma_0=-1/K\mu$是截止衰落深度，低于此值时不使用信道。注意对于第一个界（9-51），正发送功率条件($P(\gamma)\geqslant 0$)决定了截止值，而对于第二个界（9-52），正发送速率条件($k(\gamma)\geqslant 0$)决定了截止值。可用γ_0把式（9-56）表示为

$$\frac{P(\gamma)}{\bar{P}}=\begin{cases}1/\gamma_0(-K)+1/\gamma(-K) & \gamma\geqslant\gamma_0\\ 0 & \text{其他}\end{cases} \tag{9-58}$$

这是反注水法功率控制：由于K是负数，所以当信噪比在最佳截止值γ_0之上时，信噪比增加时发送功率减少。γ_0的值可用平均功率约束条件式（9-12）求得。

对于第三个界（9-53），$c_1=0.25$、$c_2=8$、$c_3=1.94$、$c_4=0$。假设目标$P_b<0.25$，则$K=-c_2/(c_4\ln(P_b/c_1))=\infty$。由式（9-49），最佳功控变成

$$\frac{P(\gamma)}{\bar{P}}=\begin{cases}\mu & k(\gamma)\geqslant 0,\ P(\gamma)\geqslant 0\\ 0 & \text{其他}\end{cases} \tag{9-59}$$

这个结果是开关传输：发送功率要么是零，要么是一个非零的常数。由式（9-44）可得相应的最佳速率控制$k(\gamma)$为

$$k(\gamma)=\begin{cases}(1/c_3)\log_2(\gamma/\gamma_0) & \gamma\geqslant\gamma_0\\ 0 & \text{其他}\end{cases} \tag{9-60}$$

式中$\gamma_0=-\ln(P_b/c_1)/c_2\mu$是截止衰落深度，低于这个值时不使用信道。和前一个界一样，截止衰落深度γ_0也是由正发送速率条件决定的。最佳功控可表示为γ_0的函数：

$$\frac{P(\gamma)}{\bar{P}}=\begin{cases}K_0/\gamma_0 & \gamma\geqslant\gamma_0\\ 0 & \text{其他}\end{cases} \tag{9-61}$$

式中$K_0=-\ln(P_b/c_1)/c_2$。γ_0的值可用平均功率约束条件求得，它满足

$$\frac{K_0}{\gamma_0}\int_{\gamma_0}^{\infty}p(\gamma)\,\mathrm{d}\gamma=1 \tag{9-62}$$

308

因此，对于MPSK误比特率的这三个界，最佳速率控制式（9-55）、式（9-57）、式（9-60）有相同的形式，但最佳功率控制式（9-54）、式（9-58）、式（9-61）的形式截然不同。图9-15画出了瑞利衰落下，目标BER为10^{-3}，平均信噪比为$\bar{\gamma}=30\text{dB}$时，式（9-54）、式（9-58）、式（9-61）所给出的注水法、反注水法和开关功率法这三种功率控制。可以注意到这几种功控的截止值基本相同。从图中同样可以看出，各种功控在低信噪比时不相同，高信噪比时是基本相同的。当$\gamma<10\text{dB}$时，差别很大，而当$\gamma\geqslant 10\text{dB}$时，它们都迅速收敛到一个相同的常数。图9-15中同时示出的γ的累计分布函数表明，γ小于10dB的概率为0.01。因此，低信噪比时功控的这种差异对频谱利用率影响不大。

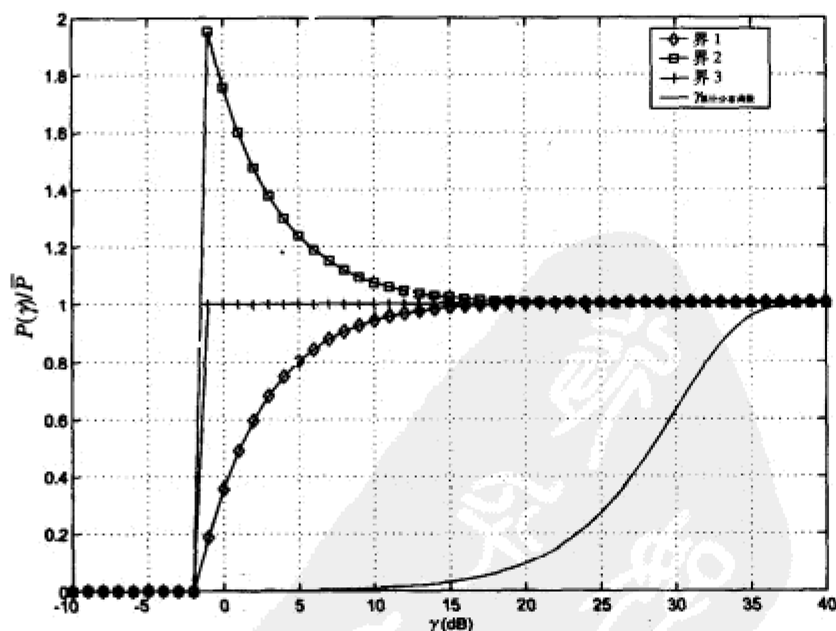


图9-15 不同MPSK误比特率界下的功率控制（瑞利衰落、$P_b=10^{-3}$、$\bar{\gamma}=30\text{dB}$）

9.4.2　离散速率的自适应

假设星座图集合为$\mathcal{M}=\{M_0=0,\cdots,M_{N-1}\}$，其中$M_0$对应无数据传输。每个星座图的速率为$k_j=\log_2 M_j(j=0,\cdots,N-1)$，其中$k_0=0$。每个速率$k_j\,(j>0)$对应一个$\gamma$的衰落区域$R_j=[\gamma_{j-1},\gamma_j)$, $j=0,\cdots,N-1$, $\gamma_{-1}=0$, $\gamma_{N-1}=\infty$。当$\gamma<\gamma_0$时，不使用信道。最佳自适应设计首先需要优化边界值$\gamma_j\,(j=0,\cdots,N-2)$。仍然假设BER近似具有式（9-43）的一般形式，则高于截止值γ_0时，满足目标误比特率的功率控制为

$$\frac{P(\gamma)}{\bar{P}}=\frac{h(k_j)}{\gamma}\quad \gamma_{j-1}\leqslant\gamma\leqslant\gamma_j \tag{9-63}$$

式中

309

$$h(k_j)=-\frac{\ln(P_b/c_1)}{c_2}(2^{c_3k_j}-c_4) \tag{9-64}$$

可用如下的拉格朗日方程求能最大化频谱利用率的边界$\gamma_0,\cdots,\gamma_{N-2}$。

$$J(\gamma_0,\gamma_1,\cdots,\gamma_{N-2})=\sum_{j=1}^{N-1}k_j\int_{\gamma_{j-1}}^{\gamma_j}p(\gamma)\,\mathrm{d}\gamma+\lambda\left[\sum_{j=1}^{N-1}\int_{\gamma_{j-1}}^{\gamma_j}\frac{h(k_j)}{\gamma}p(\gamma)\,\mathrm{d}\gamma-1\right] \tag{9-65}$$

最佳边界γ_j是下列方程组的解

$$\frac{\partial J}{\partial\gamma_j}=0\quad 0\leqslant j\leqslant N-2 \tag{9-66}$$

由此得到

$$\gamma_0=\frac{h(k_1)}{k_1}\rho \tag{9-67}$$

$$\gamma_j=\frac{h(k_{j+1})-h(k_j)}{k_{j+1}-k_j}\rho\quad 1\leqslant i\leqslant N-2 \tag{9-68}$$

式中ρ取决于平均功率的约束条件

$$\sum_{j=1}^{N-1}\int_{\gamma_{j-1}}^{\gamma_j}\frac{h(k_j)}{\gamma}p(\gamma)\,\mathrm{d}\gamma=1 \tag{9-69}$$

9.4.3　平均BER目标

现在我们考虑放松BER目标要求，不要求每个符号的BER都必须满足目标P_b，只要求平均BER低于目标的平均值$\bar{P}_b$即可。在这种情况下，除了可以调整速率和功率外，我们还可以在平均$\bar{P}_b$约束下调整瞬时BER值$P_b(\gamma)$。自由度增加有可能提高频谱利用率。定义自适应调制的平均错误率为

$$\bar{P}_b=\frac{E[\text{每次传送的误比特数}]}{E[\text{每次传送的比特数}]} \tag{9-70}$$

当比特速率$k(\gamma)$可以连续调整时，上式变为

$$\bar{P}_b = \frac{\int_0^\infty P_b(\gamma)k(\gamma)p(\gamma)\,\mathrm{d}\gamma}{\int_0^\infty k(\gamma)p(\gamma)\,\mathrm{d}\gamma} \tag{9-71}$$

当$k(\gamma)$为离散值时：

$$\bar{P}_b = \frac{\sum_{j=1}^{N-1} k_j \int_{\gamma_{j-1}}^{\gamma_j} P_b(\gamma)p(\gamma)\,\mathrm{d}\gamma}{\sum_{j=1}^{N-1} k_j \int_{\gamma_{j-1}}^{\gamma_j} p(\gamma)\,\mathrm{d}\gamma} \tag{9-72}$$

310

现在推导速率、功率和BER可连续控制时，能使频谱利用率$E[k(\gamma)]$最大化的最佳速率、功率、BER自适应方案，约束条件是平均功率$\bar{P}$和式（9-71）给出的平均BER。如同瞬时BER约束是一样，这是一个标准的条件极值问题，可用拉格朗日法来解决。只不过现在有两个约束条件：平均功率和平均BER，因此需要两个拉格朗日变量。具体的拉格朗日方程为：

$$\begin{aligned} J(k(\gamma), P(\gamma)) = & \int_0^\infty k(\gamma)p(\gamma)\,\mathrm{d}\gamma \\ & + \lambda_1\left[\int_0^\infty P_b(\gamma)k(\gamma)p(\gamma)\,\mathrm{d}\gamma - \bar{P}_b\int_0^\infty k(\gamma)p(\gamma)\,\mathrm{d}\gamma\right] \\ & + \lambda_2\left[\int_0^\infty P(\gamma)p(\gamma)\,\mathrm{d}\gamma - \bar{P}\right] \end{aligned} \tag{9-73}$$

最佳速率控制和功率控制必须满足：

$$\frac{\partial J}{\partial k(\gamma)} = 0 \quad 和 \quad \frac{\partial J}{\partial P(\gamma)} = 0 \tag{9-74}$$

并且对所有的γ，$k(\gamma)$和$P(\gamma)$必须是非负的。

假设BER具有式（9-43）所示的一般形式。定义

$$f(k(\gamma)) = 2^{c_3 k(\gamma)} - c_4 \tag{9-75}$$

将式（9-43）代入式（9-73），求解式（9-74）可得到：对于非负的$k(\gamma)$，能最大化频谱利用率的功率控制为

$$\frac{P(\gamma)}{\bar{P}} = \max\left[\frac{f(k(\gamma))}{\frac{\partial f(k(\gamma))}{\partial k(\gamma)}}\lambda_2\bar{P}(\lambda_1\bar{P}_b - 1) - \frac{f(k(\gamma))^2}{c_2\gamma\frac{\partial f(k(\gamma))}{\partial k(\gamma)}k(\gamma)}, 0\right] \tag{9-76}$$

瞬时BER为

$$P_b(\gamma) = \frac{\lambda_2\bar{P}f(k(\gamma))}{\lambda_1 c_2\gamma k(\gamma)} \tag{9-77}$$

此外，由式（9-43）、式（9-76）和式（9-77）可得到最佳速率控制$k(\gamma)$是下式的非负解或零解。

$$\frac{\lambda_1 \bar{P}_b - 1}{\frac{\partial f(k(\gamma))}{\partial k(\gamma)} \lambda_2 \bar{P}} - \frac{f(k(\gamma))}{c_2 \gamma \frac{\partial f(k(\gamma))}{\partial k(\gamma)} k(\gamma)} = \frac{1}{\gamma c_2} \ln\left[\frac{\lambda_1 c_1 c_2 \gamma k(\gamma)}{\lambda_2 \bar{P} f(k(\gamma))}\right] \tag{9-78}$$

$k(\gamma)$的值和拉格朗日乘数λ_1和λ_2必须通过数值搜索得到，并且要满足平均功率$\bar{P}$和平均约束条件。

311

在速率离散的情况下，发送速率在一个固定的集合$k_0,\cdots,k_{N-1}$内变化，其中k_0对应于没有数据传输。我们必须确定出一组区域的边界$\gamma_0,\cdots,\gamma_{N-2}$，将速率$k_j$指配到区域$[\gamma_{j-1},\gamma_j)$，其中$\gamma_{-1}=0$，$\gamma_{N-1}=\infty$。再根据平均功率和平均BER约束条件，求解能最大化频谱利用率的最佳速率、功率和瞬时BER。因为速率集合已定，速率和衰落区的指配关系已定，所以需要求解的是最佳的区域边界γ_j, $j=0,\cdots,N-2$。此条件极值问题的拉格朗日函数为：

$$\begin{aligned} J(\gamma_0,\gamma_1,\cdots,\gamma_{N-2},P(\gamma)) = &\sum_{j=1}^{N-1} k_j \int_{\gamma_{j-1}}^{\gamma_j} p(\gamma)\,\mathrm{d}\gamma \\ &+ \lambda_1\left[\sum_{j=1}^{N-1} k_j \int_{\gamma_{j-1}}^{\gamma_j} (P_b(\gamma)-\bar{P}_b)p(\gamma)\,\mathrm{d}\gamma\right] \\ &+ \lambda_2\left[\int_{\gamma_0}^{\infty} P(\gamma)p(\gamma)\,\mathrm{d}\gamma - \bar{P}\right] \end{aligned} \tag{9-79}$$

通过解下面方程中的$P(\gamma)$可得到最佳功率控制：

$$\frac{\partial J}{\partial P(\gamma)} = 0 \tag{9-80}$$

类似地，通过解下列方程中的γ_j可得到最佳边界：

$$\frac{\partial J}{\partial \gamma_j} = 0 \quad 0 \leqslant j \leqslant N-2 \tag{9-81}$$

由式（9-80）可知，最佳功率和BER控制必须满足

$$\frac{\partial P_b(\gamma)}{\partial P(\gamma)} = \frac{-\lambda_2}{k_j \lambda_1} \quad \gamma_{j-1} \leqslant \gamma \leqslant \gamma_j \tag{9-82}$$

将式（9-43）代入式（9-82），得到

$$P_b(\gamma) = \lambda \frac{f(k_j)}{\gamma k_j} \quad \gamma_{j-1} \leqslant \gamma \leqslant \gamma_j \tag{9-83}$$

式中$\lambda = \bar{P}\lambda_2 / c_2\lambda_1$。这个BER控制类似于注水法功控：信道质量好的时候，让瞬时BER低一些。现在，令式（9-43）中的BER等于式（9-83），并解出$P(\gamma)$，得到

$$P(\gamma) = P_j(\gamma) \quad \gamma_{j-1} \leqslant \gamma \leqslant \gamma_j \tag{9-84}$$

式中

$$\frac{P_j(\gamma)}{\bar{P}} = \ln\left[\frac{\lambda f(k_j)}{c_1 \gamma k_j}\right] \frac{f(k_j)}{-\gamma c_2} \quad 1 \leqslant j \leqslant N-1 \tag{9-85}$$

并且对于$\gamma<\gamma_0$有$P(\gamma)=0$。由式（9-85）可见，$P(\gamma)$在边界γ_j处是不连续的。

现在考虑最佳边界$\gamma_0,\cdots,\gamma_{N-2}$，解出式（9-81）中的$P_b(\gamma_j)$得到

$$P_b(\gamma_j)=\bar{P}_b-\frac{1}{\lambda_1}-\frac{\lambda_2}{\lambda_1}\frac{P_{j+1}(\gamma_j)-P_j(\gamma_j)}{k_{j+1}-k_j}\quad 0\leqslant j\leqslant N-2 \tag{9-86}$$

式中$k_0=0$，$P_0(\gamma)=0$。上式很难求解，但若假设$P(\gamma)$在边界处连续，则式（9-86）变为

$$P_b(\gamma_j)=\bar{P}_b-\frac{1}{\lambda_1}\quad 0\leqslant j\leqslant N-2 \tag{9-87}$$

312

在这种假设下，我们可以得到次佳的边界为：

$$\gamma_{j-1}=\frac{f(k_j)}{k_j}\rho\quad 1\leqslant j\leqslant N-1 \tag{9-88}$$

常数λ_1和ρ可通过数值搜索得到，它们必须满足平均BER约束条件（9-72）和平均功率约束条件

$$\sum_{j=1}^{N-1}\int_{\gamma_{j-1}}^{\gamma_j}\frac{P_j(\gamma)}{\bar{P}}p(\gamma)\,\mathrm{d}\gamma=1 \tag{9-89}$$

注意，$P(\gamma)$在边界处不一定连续，所以式（9-88）中的边界是次佳的，这样得到的频谱利用率也是次佳的。

图9-16画出了在瑞利衰落信道下，采用平均BER目标和瞬时BER目标时，连续速率和离散速率的自适应MQAM的平均频谱利用率。其误比特率采用式（9-7）的近似式，目标值为10^{-3}或10^{-7}。离散速率自适应采用了六种MQAM星座图（七个衰落区域），$M=\{0, 4, 16, 256, 1024, 4096\}$。从图中可以看到，在相同的瞬时BER或平均BER目标值

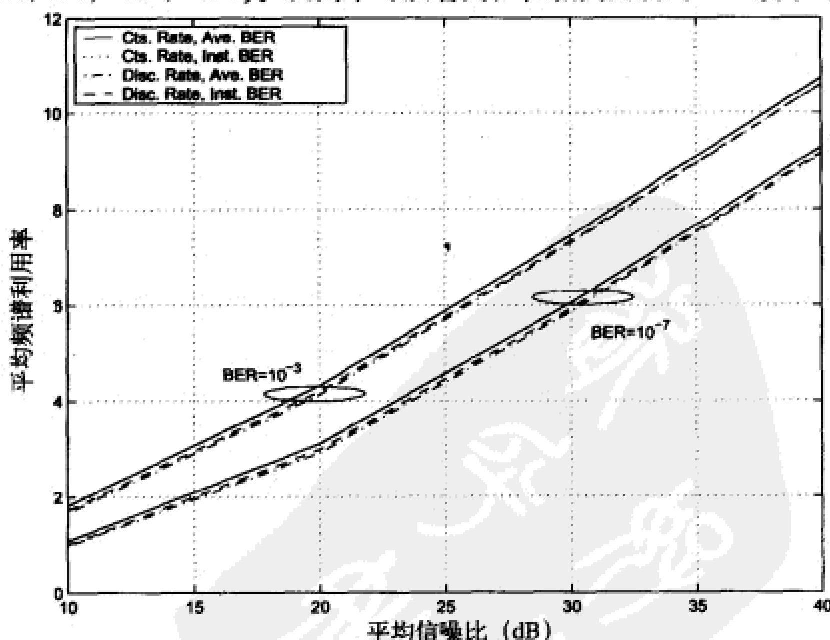


图9-16 不同自适应约束下的频谱利用率

下，四种方案的频谱利用率十分接近。对于离散速率自适应，瞬时BER目标比平均BER目标限制更严，但前者的频谱利用率却稍高，这是因为平均BER目标采用的是次佳的区域边界，这个因素使其频谱利用率有所降低。

313

9.5 结合快衰落和慢衰落的自适应技术

本节讨论结合快衰落和慢衰落（阴影衰落）的自适应技术。假设快衰落变化很快，无法做到精确测量并反馈到发送端，所以发送端只能根据慢衰落进行自适应控制。瞬时信噪比γ的分布为$p(\gamma|\bar{\gamma})$，$\bar{\gamma}$是快衰落的短期平均，它随阴影衰落缓慢变化，其分布为$p(\bar{\gamma})$，它的平均值为$\bar{\bar{\gamma}}$。发送端只能根据$\bar{\gamma}$进行自适应，因此速率$k(\bar{\gamma})$和功率$P(\bar{\gamma})$是$\bar{\gamma}$的函数。在快衰落和慢衰落时功率控制都应满足长期的平均功率约束条件：

$$\int_0^\infty P(\bar{\gamma})p(\bar{\gamma})\,\mathrm{d}\bar{\gamma} = \bar{P} \tag{9-90}$$

我们也像前面那样将瞬时BER近似为式（9-43）所示的一般形式。因为功率和速率是$\bar{\gamma}$的函数，故以$\bar{\gamma}$为条件的条件误比特率为

$$P_b(\gamma \mid \bar{\gamma}) \approx c_1 \exp\left[\frac{-c_2\gamma P(\bar{\gamma})/\bar{P}}{2^{c_3 k(\bar{\gamma})} - c_4}\right] \tag{9-91}$$

发送端不能自适应快衰落γ，所以不能要求瞬时BER达到某个给定值。但可以自适应慢衰落，故可以要求给定阴影衰落时，对快衰落进行平均后的BER满足目标值。对于给定的$\bar{\gamma}$，将$P_b(\gamma|\bar{\gamma})$按快衰落分布$p(\gamma|\bar{\gamma})$进行平均，可得到短期平均值：

$$\bar{P}_b(\bar{\gamma}) = \int_0^\infty P_b(\gamma \mid \bar{\gamma})p(\gamma \mid \bar{\gamma})\,\mathrm{d}\gamma \tag{9-92}$$

将式（9-91）代入式（9-92）并假设快衰落为瑞利衰落，结果成为：

$$\bar{P}_b(\bar{\gamma}) = \frac{1}{\bar{\gamma}}\int_0^\infty c_1 \exp\left[\frac{-c_2\gamma P(\bar{\gamma})/\bar{P}}{2^{c_3 k(\bar{\gamma})} - c_4} - \frac{\gamma}{\bar{\gamma}}\right]\mathrm{d}\gamma = \frac{c_1}{\dfrac{1.5\bar{\gamma}P(\bar{\gamma})/\bar{P}}{2^{c_3 k(\bar{\gamma})} - c_4} + 1} \tag{9-93}$$

例如，对于MQAM调制，利用式（9-7）的BER界，式（9-93）成为

$$\bar{P}_b(\bar{\gamma}) = \frac{0.2}{\dfrac{1.5\bar{\gamma}P(\bar{\gamma})/\bar{P}}{2^{k(\bar{\gamma})} - 1} + 1} \tag{9-94}$$

对式（9-93）求逆，得到速率控制$k(\bar{\gamma})$作为目标平均误比特率$\bar{P}_b$和功率控制$P(\bar{\gamma})$的函数：

$$k(\bar{\gamma}) = \frac{1}{c_3}\log_2\left(c_4 + \frac{K\bar{\gamma}P(\bar{\gamma})}{\bar{P}}\right) \tag{9-95}$$

式中

314

$$K = \frac{c_2}{c_1/\bar{P}_b - 1} \tag{9-96}$$

它只取决于目标平均误比特率，目标值越小则K也越小。最大化频谱利用率就是最大化：

$$E[k(\bar{\gamma})]=\int_0^{\infty}\frac{1}{c_3}\log_2\left(c_4+\frac{K\bar{\gamma}P(\bar{\gamma})}{\bar{P}}\right)p(\bar{\gamma})\,\mathrm{d}\bar{\gamma} \tag{9-97}$$

约束条件是由式（9-90）给出的平均功率。

假设$c_4>0$，那么这个条件极值问题和式（9-11）完全相同，只是用慢衰落$\bar{\gamma}$代替了快衰落γ。因而最佳功控也是（9-13）这样的注水法，为

$$\frac{P(\bar{\gamma})}{\bar{P}}=\begin{cases}1/\bar{\gamma}_0-c_4/\bar{\gamma}K & \bar{\gamma}\geqslant c_4\bar{\gamma}_0/K\\ 0 & \bar{\gamma}<c_4\bar{\gamma}_0/K\end{cases} \tag{9-98}$$

当$\bar{\gamma}<c_4\bar{\gamma}_0/K$时不使用信道。$\bar{\gamma}_0$的值由平均功率约束条件决定。将式（9-98）代入式（9-95）得到速率控制为

$$k(\bar{\gamma})=\frac{1}{c_3}\log_2\left(\frac{K\bar{\gamma}}{\bar{\gamma}_0}\right) \tag{9-99}$$

相应的平均频谱利用率为

$$\frac{R}{B}=\int_{c_4\gamma_0/K}^{\infty}\log_2\left(\frac{K\bar{\gamma}}{\bar{\gamma}_0}\right)p(\bar{\gamma})\,\mathrm{d}\bar{\gamma} \tag{9-100}$$

可见，对于结合快衰落和慢衰落的信道，如果功率和速率只根据慢衰落进行自适应，并且式（9-43）中的$c_4>0$，那么满足平均BER约束的最佳功控是相对于慢衰落的注水法功控。

虽然我们的推导中假设了快衰落为瑞利衰落，但文献[34]中指出，只要式（9-43）中的$c_4>0$，对于任意的快衰落分布，最佳功率控制和速率控制都具有相同的形式。这是因为假设$c_4>0$后，正功率约束条件决定了不使用信道的截止值。如同我们在9.4.1节中所见到的，当$c_4<0$时，正速率约束条件决定了截止值，此时的最佳功率控制是反注水法，而在$c_4=0$时则为开关功率控制。

习题

9.1　假设瑞利衰落、8-PSK调制，求平均误比特率为$\bar{P}_b=10^{-3}$所需要的平均信噪比及相应的频谱利用率。假设码元间隔$T_s=1/B$。

9.2　假设瑞利衰落的平均信噪比为20dB，采用截断式信道反转。问中断率为0.1时的σ为多少？若正常通信期间的$P_b\approx10^{-3}$，求所对应的最大MQAM星座大小。

315

9.3　假设瑞利信道，其$\bar{\gamma}=20$dB，采用QPSK调制，如欲无中断时的$P_b=10^{-3}$，求功率控制方案以及相应的中断率。

9.4　某自适应速率的MQAM系统只有$M=4$和$M=16$两种星座图。目标P_b近似为10^{-3}，不能满足这个目标时不传数据。

（a）利用式（9-7）中的BER界，求满足目标BER时三种传输方式（不传输、4-QAM、

16-QAM）各自对应的γ值范围。

（b）假设瑞利衰落，$\bar{\gamma}=20\text{dB}$，求此自适应速率系统的平均数据速率。

（c）假设$0\leqslant\gamma\leqslant\gamma_0$时用BPSK进行传输，请用式（9-6）求出BPSK的平均BER。

9.5 某自适应调制编码系统有三种调制方式：BPSK、QPSK和8-PSK，有三种分组码编码，码率分别为1/2、1/3、1/4。假设对任意调制类型，第一种编码的编码增益大致为3dB，第二种为4dB，第三种为5dB。对所有可能的信噪比$0\leqslant\gamma\leqslant\infty$，求出能满足目标误比特率$10^{-3}$的速率最大的调制编码组合（计算中可以使用任何合理的BER近似式，将编码增益直接加到信噪比上）。假设任何调制编码组合都不能达到目标BER时，不传输数据。对于平均信噪比为20dB的瑞利衰落，求该系统的平均数据速率。

9.6 证明在功率约束条件式（9-12）下，式（9-13）给出的注水功控可以使频带利用率式（9-11）最大化。证明时先建立拉格朗日方程，对其微分，再得到最优功率分配。同时证明，采用这种功率分配时的速率控制为式（9-16）。

9.7 本题比较自适应系统和非自适应系统的频谱利用率

316

（a）利用式（9-7）中MQAM的BER紧界，求瑞利衰落下的平均BER作为M和$\bar{\gamma}$的函数。

（b）假设瑞利衰落，$\bar{\gamma}=20\text{dB}$。利用（a）中的结果求出目标平均BER为10^{-3}所对应的最大星座。

（c）将（b）中得到的频谱利用率和图9-3中相同参数的自适应调制的频谱利用率相比较。自适应系统和非自适应系统的频谱利用率相差多少？

9.8 考虑平均信噪比为20dB的瑞利衰落信道，假设目标误比特率为10^{-4}。

（a）对于速率可变功率可变的MQAM，求最佳的速率控制和功率控制，以及截止值γ_0/K。

（b）对于（a）中的自适应方案，求平均频谱利用率。

（c）比较（b）的频谱利用率和γ_0经过优化了的截断式信道反转的频谱利用率。

9.9 考虑一个四状态的离散时变AWGN信道。假设发送功率固定为$\bar{P}$时，这四种信道状态下的接收信噪比分别为$\gamma_1=5\text{dB}$、$\gamma_2=10\text{dB}$、$\gamma_3=15\text{dB}$、$\gamma_4=20\text{dB}$。各信道状态的出现概率为$p(\gamma_1)=0.4$、$p(\gamma_2)=p(\gamma_3)=p(\gamma_4)=0.2$。假设目标误比特率为$10^{-3}$。

（a）对于连续速率的自适应MQAM，求最佳的功率控制和速率控制。

（b）求这种最佳自适应方案的平均频谱利用率。

（c）求此信道下的截断式信道反转功率控制，以及所能达到的最大数据速率。

9.10 考虑平均信噪比为20dB的瑞利衰落信道，要求的误比特率为10^{-3}。假设星座图的大小限制为0、2、4、16、64及256，求截断式信道反转下的频谱利用率。

9.11 考虑平均信噪比为20dB、多普勒频率为80Hz的瑞利衰落信道，要求的误比特率为10^{-3}。

（a）假设自适应MQAM的星座图大小限制为0、2、4、16和64。若$\gamma_K^*=0.1$，求出每一种星座图所对应的衰落区R_j。并求该自适应调制的平均频谱利用率以及每个衰落区R_j的平均出现时间。如果码元间隔$T_s=B^{-1}$，问每一种星座大约会连续发送多少个符号？

（b）利用式（9-33）求出该自适应方案的精确误比特率，它和目标误比特率有何差别？

9.12　考虑平均信噪比为20dB、多普勒频率为80Hz的瑞利衰落信道，信号带宽为30kHz，要求的误比特率为10^{-3}。

(a) 假设速率可变功率可变的MQAM系统中估计误差$\varepsilon=\hat{\gamma}/\gamma$均匀分布在0.5和1.5之间，求该系统的平均误比特率。

(b) 假设速率可变功率可变MQAM系统中发送端的信噪比估计$\hat{\gamma}$同时存在延时和随机误差：$\hat{\gamma}(t)=\gamma(t-\tau)+\gamma_\varepsilon(t)$，为了计算平均误码率，我们需要什么样的联合分布？

9.13　考虑平均信噪比为20dB的瑞利信道以及编码增益为3dB的自适应网格码MQAM系统。假设目标误比特率为10^{-3}，求最佳的功率控制和速率控制，以及相应的平均频谱利用率。

317

9.14　根据第6章，非矩形MQAM的P_b界为

$$P_b \approx \frac{4}{\log_2 M} Q\left(\sqrt{\frac{3\gamma}{(M-1)}}\right)$$

求$M=8$时，用式（9-43）的一般形式近似此BER时的c_1、c_2、c_3、c_4的值。可以使用任何曲线拟合方法。画出$0\leqslant\gamma\leqslant 30$dB时的两种BER曲线。

9.15　证明：当功率约束为$\bar{P}$，$k(\gamma)$由式（9-44）给出时，式（9-47）给出的功控能最大化平均频谱利用率$E[k(\gamma)]$。

9.16　本题在BER界（9-51）、（9-52）和（9-53）的基础上研究最佳的自适应MPSK调制。假设瑞利衰落（γ为指数分布，$\bar{\gamma}=30$dB），目标误比特率为$P_b=10^{-7}$。

(a) 截止衰落深度γ_0必须满足

$$\int_{\gamma_0/K}^{\infty}\left(\frac{1}{\gamma_0}-\frac{1}{\gamma K}\right)p(\gamma)d\gamma \leqslant 1$$

其中的K由式（9-48）给出，求采用不同BER界所得到的截止值γ_0。

(b) 画出这三种界下$P(\gamma)/\bar{P}$和$k(\gamma)$的曲线，γ的范围是0~30dB。并说明，在不同界下，γ_0是由正功率决定的还是由正速率决定的。

(c) 在低信噪比时，不同界对应的功率控制有何不同？高信噪比时又如何？

9.17　证明：对于离散速率控制的一般M进制调制，能满足瞬时BER目标的功率控制由式（9-63）给出。同时证明，利用拉格朗日函数式（9-65）得到的最大频谱利用率对应的边界值为式（9-67）和式（9-68）。

9.18　请证明，对于平均BER约束的一般M进制调制，从拉格朗日函数式（9-80）导出最佳功率控制和BER控制必然满足式（9-82）。再从式（9-82）导出BER控制式（9-83），从式（9-83）导出功率控制式（9-84）和（9-85）。最后再从式（9-81）证明最佳区域边界必然满足式（9-86）。

9.19　假设自适应MPSK的星座图限制为：没有数据传输，或者$M=2$、4、8、16。假设其误比特率可用式（9-51）近似。求出最佳的离散速率连续功率自适应方案，在$0\leqslant\gamma\leqslant$ 30dB范围内画出其曲线，并求出平均频谱利用率。假设瑞利信道，$\bar{\gamma}=20$dB，目标误比特率为10^{-4}。

9.20　考虑与上题相同的离散速率自适应MPSK，但BER约束是平均值为10^{-4}，而不是

瞬时值。假设瑞利信道，$\bar{\gamma}=20\text{dB}$。求最佳离散速率连续功率自适应方案及其相应的平均频谱利用率。

318

9.21 考虑结合了快速瑞利衰落和慢速对数正态阴影衰落的信道，其分贝平均的信噪比是$\mu_{\psi_{dB}}=20\text{dB}$（同时对快衰落和慢衰落平均），$\sigma_{\psi_{dB}}=8\text{dB}$。假设自适应MPSK系统只适应阴影衰落，目标的平均误比特率为10^{-3}。请利用式（9-51）的BER近似式，求出能最大化平均频谱利用率，并能满足平均误比特率目标值且为慢衰落$\bar{\gamma}$的函数的最佳功率控制和速率控制，并求出相应的平均频谱利用率。

9.22 我们在9.5节中讨论了结合瑞利衰落和阴影衰落时，能满足目标的平均误比特率，并能最大化平均频谱利用率的速率和功率自适应。推导过程采用了$c_4>0$的BER近似式(9-43)。现在请按$c_4<0$推导相应的功率控制和速率控制。提示：这个推导过程类似于使用第二个MPSK误比特率界的连续速率自适应，结果也同样是反注水法功率控制。

9.23 仍然考虑上题中结合瑞利衰落和阴影衰落的信道下，满足目标误比特率时，能最大化平均频谱利用率的速率自适应和功率自适应。但本题假设式（9-43）中的$c_4=0$。请按$c_4=0$推导相应的功率控制和速率控制。提示：这个推导过程类似于9.4.1节中用第三个MPSK误比特率界的情形，结果也是开关功率控制。

参考文献

[1] J. F. Hayes, "Adaptive feedback communications," *IEEE Trans. Commun. Tech.*, pp. 29–34, February 1968.

[2] J. K. Cavers, "Variable-rate transmission for Rayleigh fading channels," *IEEE Trans. Commun.*, pp. 15–22, February 1972.

[3] S. Otsuki, S. Sampei, and N. Morinaga, "Square-QAM adaptive modulation/TDMA/TDD systems using modulation level estimation with Walsh function," *Elec. Lett.*, pp. 169–71, February 1995.

[4] W. T. Webb and R. Steele, "Variable rate QAM for mobile radio," *IEEE Trans. Commun.*, pp. 2223–30, July 1995.

[5] Y. Kamio, S. Sampei, H. Sasaoka, and N. Morinaga, "Performance of modulation-level-controlled adaptive-modulation under limited transmission delay time for land mobile communications," *Proc. IEEE Veh. Tech. Conf.*, pp. 221–5, July 1995.

[6] B. Vucetic, "An adaptive coding scheme for time-varying channels," *IEEE Trans. Commun.*, pp. 653–63, May 1991.

[7] M. Rice and S. B. Wicker, "Adaptive error control for slowly varying channels," *IEEE Trans. Commun.*, pp. 917–26, February–April 1994.

[8] S. M. Alamouti and S. Kallel, "Adaptive trellis-coded multiple-phased-shift keying for Rayleigh fading channels," *IEEE Trans. Commun.*, pp. 2305–14, June 1994.

[9] T. Ue, S. Sampei, and N. Morinaga, "Symbol rate and modulation level controlled adaptive modulation/TDMA/TDD for personal communication systems," *Proc. IEEE Veh. Tech. Conf.*, pp. 306–10, July 1995.

[10] H. Matsuoka, S. Sampei, N. Morinaga, and Y. Kamio, "Symbol rate and modulation level controlled adaptive modulation/TDMA/TDD for personal communication systems," *Proc. IEEE Veh. Tech. Conf.*, pp. 487–91, April 1996.

[11] S. Sampei, N. Morinaga, and Y. Kamio, "Adaptive modulation/TDMA with a BDDFE for 2 Mbit/s multi-media wireless communication systems," *Proc. IEEE Veh. Tech. Conf.*, pp. 311–15, July 1995.

319

[12] S. T. Chung and A. J. Goldsmith, "Degrees of freedom in adaptive modulation: A unified view," *IEEE Trans. Commun.*, pp. 1561–71, September 2001.

[13] A. Furuskar, S. Mazur, F. Muller, and H. Olofsson, "EDGE: Enhanced data rates for GSM and TDMA/136 evolution," *IEEE Wireless Commun. Mag.*, pp. 56–66, June 1999.

[14] A. Ghosh, L. Jalloul, B. Love, M. Cudak, and B. Classon, "Air-interface for 1XTREME/1xEV-DV," *Proc. IEEE Veh. Tech. Conf.*, pp. 2474–8, May 2001.

[15] S. Nanda, K. Balachandran, and S. Kumar, "Adaptation techniques in wireless packet data services," *IEEE Commun. Mag.*, pp. 54–64, January 2000.

[16] H. Sari, "Trends and challenges in broadband wireless access," *Proc. Sympos. Commun. Veh. Tech.*, pp. 210–14, October 2000.

[17] K. M. Kamath and D. L. Goeckel, "Adaptive-modulation schemes for minimum outage probability in wireless systems," *IEEE Trans. Commun.*, pp. 1632–5, October 2004.

[18] J. Hagenauer, "Rate-compatible punctured convolutional codes (RCPC codes) and their applications," *IEEE Trans. Commun.*, pp. 389–400, April 1988.

[19] G. J. Foschini and J. Salz, "Digital communications over fading radio channels," *Bell System Tech. J.*, pp. 429–56, February 1983.

[20] G. D. Forney, Jr., R. G. Gallager, G. R. Lang, F. M. Longstaff, and S. U. Quereshi, "Efficient modulation for band-limited channels," *IEEE J. Sel. Areas Commun.*, pp. 632–47, September 1984.

[21] J. G. Proakis, *Digital Communications*, 4th ed., McGraw-Hill, New York, 2001.

[22] M. Filip and E. Vilar, "Implementation of adaptive modulation as a fade countermeasure," *Internat. J. Sat. Commun.*, pp. 181–91, 1994.

[23] C. C. Tan and N. C. Beaulieu, "On first-order Markov modeling for the Rayleigh fading channel," *IEEE Trans. Commun.*, pp. 2032–40, December 2000.

[24] L. Kleinrock, *Queueing Systems*, vol. I: *Theory*, Wiley, New York, 1975.

[25] W. C. Jakes Jr., *Microwave Mobile Communications*, Wiley, New York, 1974.

[26] H. S. Wang and P.-C. Chang, "On verifying the first-order Markov assumption for a Rayleigh fading channel model," *IEEE Trans. Veh. Tech.*, pp. 353–7, May 1996.

[27] X. Tang, M.-S. Alouini, and A. Goldsmith, "Effect of channel estimation error on M-QAM BER performance in Rayleigh fading," *IEEE Trans. Commun.*, pp. 1856–64, December 1999.

[28] J. K. Cavers, "An analysis of pilot symbol assisted modulation for Rayleigh fading channels," *IEEE Trans. Veh. Tech.*, pp. 686–93, November 1991.

[29] A. J. Goldsmith and L. J. Greenstein, "Effect of average power estimation error on adaptive MQAM modulation," *Proc. IEEE Internat. Conf. Commun.*, pp. 1105–9, June 1997.

[30] D. L. Goeckel, "Adaptive coding for time-varying channels using outdated fading estimates," *IEEE Trans. Commun.*, pp. 844–55, June 1999.

[31] A. Duel-Hallen, S. Hu, and H. Hallen, "Long-range prediction of fading signals," *IEEE Signal Proc. Mag.*, pp. 62–75, May 2000.

[32] M.-S. Alouini and A. J. Goldsmith, "Adaptive modulation over Nakagami fading channels," *Kluwer J. Wireless Pers. Commun.*, pp. 119–43, May 2000.

[33] S.-G. Chua and A. J. Goldsmith, "Adaptive coded modulation for fading channels," *IEEE Trans. Commun.*, pp. 595–602, May 1998.

[34] S. Vishwanath, S. A. Jafar, and A. J. Goldsmith, "Adaptive resource allocation in composite fading environments," *Proc. IEEE Globecom Conf.*, pp. 1312–16, November 2001.

320

第10章　多天线和空时通信

本章主要考虑发送端和接收端均有多个天线的系统，通常称为多人多出（multiple-input multiple-output，MIMO）系统。多天线系统可以通过复用来提高数据速率，也可以通过分集来提高性能。第7章中已经讨论了分集，MIMO系统中的发送天线和接收天线都能进行分集。复用是利用信道增益矩阵的结构来获得独立的信号通路，从而可以在这些通路上发送独立的数据。确切地说，MIMO最初引起人们关注是因为Winters[1]、Foschini[2]、Gans[3]和Telatar[4-6]等人预言说，发送端和接收端配置多个天线能显著提高频谱效率。获得这些频谱效率增益一般需要接收端甚至发送端精确已知信道信息。除了频谱效率增益外，利用智能天线技术还可以有效地减少ISI和来自其他用户的干扰。通过MIMO技术获得性能增益的代价是增加安装多个天线所需要的空间和功耗（这个问题对于小型的手持终端尤其突出），再就是增加了多维信号处理的复杂度。本章将研究多天线的各种不同用法，并找到各自的性能优势。文中要用到矩阵理论中的一些主要结果，附录C对此进行了简单回顾。

321

10.1　窄带MIMO模型

这一节主要考虑窄带MIMO信道。图10-1所示为窄带点对点通信系统，发送端天线数为M_t，接收端天线数为M_r。此系统可由下面的离散时间模型来表示：

$$\begin{bmatrix} y_1 \\ \vdots \\ y_{M_r} \end{bmatrix} = \begin{bmatrix} h_{11} & \cdots & h_{1M_t} \\ \vdots & \ddots & \vdots \\ h_{M_r1} & \cdots & h_{M_rM_t} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_{M_t} \end{bmatrix} + \begin{bmatrix} n_1 \\ \vdots \\ n_{M_r} \end{bmatrix}$$

也可简单表示为$\boldsymbol{y}=\boldsymbol{H}\boldsymbol{x}+\boldsymbol{n}$。$\boldsymbol{x}$代表$M_t$维发送符号，$\boldsymbol{n}$代表$M_r$维噪声向量，$\boldsymbol{H}$为$M_r \times M_t$的信道增益矩阵，其中的$h_{ij}$代表从发送天线$j$到接收天线$i$的信道增益。假设信道带宽为$B$，复高斯噪声的均值为零、协方差矩阵为$\sigma^2\boldsymbol{I}_{M_r}$，其中 $\sigma^2 \triangleq E[n_i^2] = N_0/2$ ，$N_0/2$是信道噪声的

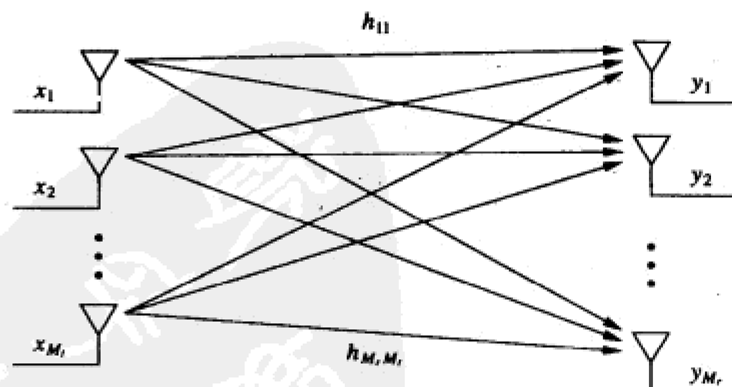


图10-1　MIMO系统

功率谱密度。简单起见，我们可以把发送功率为P，噪声功率为σ^2的情形等价成发送功率为$P/\sigma^2=\rho$，将ρ理解为归一化信道增益条件下每接收天线的平均信噪比(SNR)。此时输入符号满足

$$\sum_{i=1}^{M_t} E[x_i x_i^*] = \rho \tag{10-1}$$

或者相当于Tr ($\boldsymbol{R}_x$)=ρ，其中Tr ($\boldsymbol{R}_x$)代表输入协方差矩阵 $\boldsymbol{R}_x=\boldsymbol{E}[\boldsymbol{x}\boldsymbol{x}^H]$的迹。

关于发送端和接收端是否已知信道增益矩阵$\boldsymbol{H}$，我们可以做出不同的假设：发送端信道边信息（channel side information at the transmitter，CSIT）和接收端信道边信息（channel side information at the receiver，CSIR）。静态信道一般可以假设CSIR，因为其信道增益容易通过发送导频序列进行信道估计来获得。有关MIMO信道估计的技术详见文献[6]的3.9节。如果存在反馈信道的话，接收端的CSIR可以反馈给发送端，从而得到CSIT：在双向通信系统中，也有可能无需反馈信道，利用电波传播的互易性来得到CSIT。当发送端和接收端都不知道信道状态时，一般可以假设信道增益矩阵具有某种分布模型。最常用的模型是零均值空间白（zero-mean spatially white，ZMSW）模型。在这个模型中，$\boldsymbol{H}$的元素是独立同分布的零均值、单位方差的复循环对称高斯随机变量[1]。除非另有说明，我们始终采用ZMSW模型。也有可能矩阵元素是复循环对称高斯随机变量，但均值不为零，或者协方差矩阵不是单位阵。一般来说，对信道信息和$\boldsymbol{H}$的分布假设不同时，相应的信道容量和空时编码方法一般也不同。

322

接收信号的最佳解码是最大似然解调。如果每根发送天线上调制的符号是从大小为$|\chi|$ 的符号集中选出的，那么由于发送信号在接收端互相耦合，最大似然解调需要搜索所有输入向量，每个输入向量有M_t 个符号，所有可能的向量有 $|\chi|^{M_t}$ 个。对于一般的信道矩阵来说，如果发送端不知道$\boldsymbol{H}$，最大似然解调的复杂度很难降低，即使对较少的发送天线数，译码复杂度也非常大。但在下节中将会看到，如果发送端知道信道信息，那么接收端译码的复杂度可以大大降低。

10.2 MIMO信道的并行分解

我们在第7章提到，多个发送或接收天线可以带来分集增益。发送端和接收端都配有多个天线时，还可以获得另外一种增益，称作*复用增益*（multiplexing gain）。MIMO系统的复用增益来源于MIMO信道可以分解成R个并行的独立信道。在这些独立信道上传输多路数据，数据速率就可以比单天线系统提高R倍，这个提高的倍数就是复用增益。这一节我们将介绍怎样从MIMO系统中获得这些独立信道。

考虑一个$M_r \times M_t$ 的MIMO信道，发送端和接收端都已知信道增益矩阵$\boldsymbol{H}$。令R_H表

1. 如果对于任意的$\theta\in[0,2\pi]$，复随机变量矢量$\boldsymbol{x}$的分布与$e^{j\theta}\boldsymbol{x}$的分布相同，则称$\boldsymbol{x}$为循环对称的。因此$\boldsymbol{x}$的实部和虚部必然是独立同分布的。若$\boldsymbol{x}$为循环对称，取$\theta=\pi$可得$\boldsymbol{E}[x]=0$，取$\theta=\pi/2$可得$\boldsymbol{E}[\boldsymbol{x}\boldsymbol{x}^T]=0$，即$\boldsymbol{x}$是常态的随机向量（proper random vector）。反之，若零均值的复随机向量$\boldsymbol{x}$为常态，则它是循环对称的。

示H的秩。由附录C可知，对任意的H，可进行如下的奇异值分解（singular value decomposition，SVD）：

$$H=U\Sigma V^H \tag{10-2}$$

其中$M_r\times M_r$的矩阵U和$M_t\times M_t$的矩阵V是酉阵[1]，$M_r\times M_t$的矩阵Σ是由H的奇异值$\{\sigma_i\}$构成的对角阵。这些奇异值中有R_H个不为零，并且$\sigma_i=\sqrt{\lambda_i}$，其中$\lambda_i$为$HH^H$的第$i$大特征值。因为矩阵$H$的秩$R_H$不可能超过它的行数或列数，所以$R_H\leqslant\min(M_t, M_r)$。满秩的情形称为**富散射环境**（rich scattering environment），此时$R_H=\min(M_t, M_r)$。其他情况可能是低秩的，若某个信道中H的元素高度相关，其秩可能会降为1。

用**发送预编码**（transmit precoding）和**接收成形**（receiver shaping）对信道的输入输出x和y分别进行变换，就可以实现信道的并行分解。发送预编码将输入向量x经线性变换$x=V\tilde{x}$后作为天线的输入，接收成形将信道的输出y乘以U^H，如图10-2所示。

323

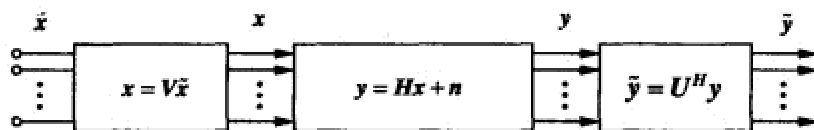


图10-2　发送预编码和接收成形

发送预编码和接收成形将MIMO信道变换成R_H个并行的单入单出（single-input single-output，SISO）信道，其输入为$\tilde{x}$，输出为$\tilde{y}$。这一点可从奇异值分解得到：

$$\begin{aligned}\tilde{y}&=U^H(Hx+n)\\&=U^H(U\Sigma V^Hx+n)\\&=U^H(U\Sigma V^HV\tilde{x}+n)\\&=U^HU\Sigma V^HV\tilde{x}+U^Hn\\&=\Sigma\tilde{x}+\tilde{n}\end{aligned}$$

其中$\tilde{\mathbf{n}}=U^Hn$，Σ是由H的奇异值构成的对角阵。注意到乘上酉阵不改变噪声的分布，即n和$\tilde{\mathbf{n}}$是同分布的。这样，发送预编码和接收成形将MIMO信道变换成了R_H个独立的并行信道，第i个信道的输入为$\tilde{x}_i$、输出为$\tilde{y}_i$、噪声为$\tilde{n}_i$、信道增益为σ_i。注意σ_i之间是有关联的，它们都是H的函数。不过由于这些并行信道并不相互干扰，所以我们说这些信道是通过总发送功率联系在一起的一组独立信道。并行分解如图10-3所示。并行的信道互不干扰，使得最大似然解调的复杂度随R_H线性增长。此外，通过在这些信道上发送独立数据，MIMO的数据速率将是单天线系统的R_H倍，即复用增益为R_H。但需注意，每个信道的性能与σ_i有关。下一节将结合MIMO信道的香农容量详细分析复用增益。

324

1. 说U和V是酉阵的意思就是$U^HU=I_{M_r}$、$V^HV=I_{M_t}$。

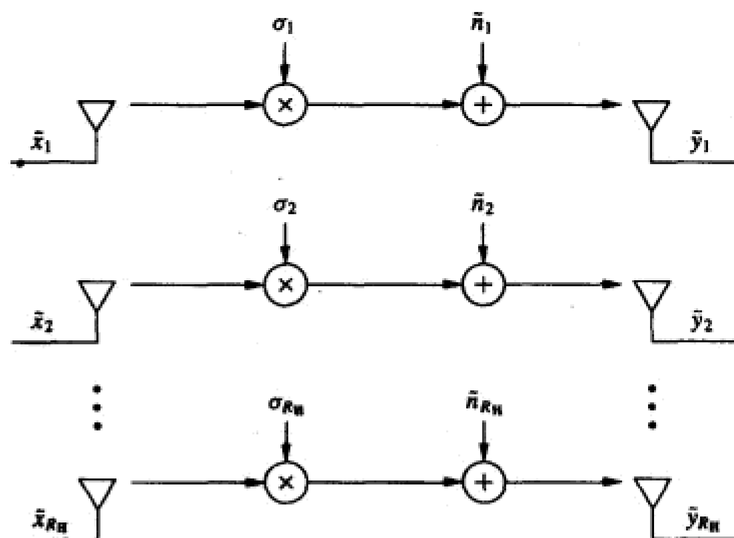


图10-3 MIMO信道的并行分解

例10.1：某MIMO信道的信道增益矩阵为

$$H=\begin{bmatrix}0.1 & 0.3 & 0.7\\ 0.5 & 0.4 & 0.1\\ 0.2 & 0.6 & 0.8\end{bmatrix} \tag{10-3}$$

求等价的并行信道模型。

解：H的奇异值分解为

$$H=\begin{bmatrix}-0.555 & 0.3764 & -0.7418\\ -0.3338 & -0.9176 & -0.2158\\ -0.7619 & -0.1278 & -0.6349\end{bmatrix}\begin{bmatrix}1.3333 & 0 & 0\\ 0 & 0.5129 & 0\\ 0 & 0 & 0.0965\end{bmatrix}\times\begin{bmatrix}-0.2811 & -0.7713 & -0.5710\\ -0.5679 & -0.3459 & -0.7469\\ -0.7736 & -0.5342 & -0.3408\end{bmatrix} \tag{10-4}$$

结果有3个非零奇异值，故R_H=3，可以分解为3个并行信道，信道增益分别为σ_1=1.3333、σ_2=0.5129、σ_3=0.0965。可以看到，信道增益逐渐减小，第三个信道的增益非常小，这个信道要么差错率很高，要么信道容量很小。

10.3 MIMO信道的容量

本节讨论MIMO信道的香农容量，它是能够以任意小差错率传输的最大数据率。中断容量则定义为能使中断率不超过某个值的最大数据率。信道容量的大小和收发两端是否已知信道增益矩阵或其分布有关。本节先给出不同信道信息假设下静态信道的容量，它是其后所讨论的衰落信道容量的基础。

10.3.1 静态信道

MIMO信道的容量是SISO信道的互信息公式（4-3）在矩阵信道下的扩展。静态信道下，接收端可以容易地对$\boldsymbol{H}$做出很好的估计，因此本节假设有CSIR。在此假设下，信道容量由信道输入向量$\boldsymbol{x}$和输出向量$\boldsymbol{y}$之间的互信息确定：

325

$$C=\max_{p(\mathbf{x})} I(X;Y)=\max_{p(\mathbf{x})}[H(Y)-H(Y\mid X)] \tag{10-5}$$

其中$H(Y)$是$\boldsymbol{y}$的熵，$H(Y|X)$是$\boldsymbol{y}|\boldsymbol{x}$的熵，熵的定义见4.1节[1]。由熵的定义可知，$H(Y|X)=H(\boldsymbol{n})$，其中$H(\boldsymbol{n})$为噪声的熵。噪声的熵独立于信道输入，所以最大化互信息就是要最大化$\boldsymbol{y}$的熵。

若给定输入向量$\boldsymbol{x}$的协方差矩阵为$\boldsymbol{R}_x$，那么MIMO信道输出$\boldsymbol{y}$的协方差矩阵$\boldsymbol{R}_y$为

$$\boldsymbol{R}_y=E[\boldsymbol{y}\boldsymbol{y}^H]=\boldsymbol{H}\boldsymbol{R}_x\boldsymbol{H}^H+\boldsymbol{I}_{M_r} \tag{10-6}$$

可以证明，给定协方差矩阵为$\boldsymbol{R}_y$的所有随机向量中，零均值循环对称复高斯（zero-mean, circularly symmetric complex Gaussian，ZMCSCG）随机向量的熵最大[2]。而仅当输入向量$\boldsymbol{x}$是ZMCSCG时，$\boldsymbol{y}$才是ZMCSCG，所以ZMCSCG是式（10-5）中$\boldsymbol{x}$的最佳分布，功率约束条件是Tr（$\boldsymbol{R}_x$）=ρ。于是，我们有 $H(Y)=B\log_2\det[\pi e\boldsymbol{R}_y]$, $H(\boldsymbol{n})=B\log_2\det[\pi e\boldsymbol{I}_{M_r}]$,从而互信息为

$$I(X;Y)=B\log_2\det[\boldsymbol{I}_{M_r}+\boldsymbol{H}\boldsymbol{R}_x\boldsymbol{H}^H] \tag{10-7}$$

文献[3]、[5]针对多天线系统的互信息导出了此式，它也出现在早期关于MIMO的文献[7~8]中，以及关于ISI信道矩阵模型的文献[9~10]中。

MIMO信道的容量就是所有满足功率约束条件的输入协方差矩阵$\boldsymbol{R}_x$中，使（10-7）的互信息最大，即

$$C=\max_{\boldsymbol{R}_x:\mathrm{Tr}(\boldsymbol{R}_x)=\rho} B\log_2\det[\boldsymbol{I}_{M_r}+\boldsymbol{H}\boldsymbol{R}_x\boldsymbol{H}^H] \tag{10-8}$$

其中det[$\boldsymbol{A}$]是矩阵$\boldsymbol{A}$的行列式。很明显，最优的$\boldsymbol{R}_x$取决于发送端是否知道$\boldsymbol{H}$。下面我们根据不同的发送端CSI假设来考虑这个最优问题。

1. 发送端已知信道：注水法

对于给定信道矩阵为$\boldsymbol{H}$的MIMO信道，当发送端和接收端均都已知$\boldsymbol{H}$时，10.2节所述的MIMO分解可以简化信道容量的分析。具体而言，信道容量等于总发送功率在各个信道之间最优分配后，各个独立并行信道的容量之和。最优功率分配就是通过优化协方差矩阵使（10-8）最大化的结果。将奇异值分解（10-2）代入（10-8），利用酉阵性质可以得到收发两端都已知信道时MIMO信道的容量为

326

$$C=\max_{\rho_i:\Sigma_i\rho_i\le\rho}\sum_{i=1}^{R_H}B\log_2(1+\sigma_i^2\rho_i) \tag{10-9}$$

其中R_H是$\boldsymbol{H}$的非零奇异值的个数。由于MIMO信道可分解成R_H个并行信道，因此称其自

1. 4.1节定义了随机标量的熵，此定义同样适用于随机矢量。

由度（degree of freedom）为R_H。由于$\rho=P/\sigma^2$，式（10-9）所示的容量也可以根据第i个并行信道的功率P_i表示为

$$C=\max_{P_i:\sum_i P_i<P}\sum_{i=1}^{R_H}B\log_2\left(1+\frac{\sigma_i^2P_i}{\sigma^2}\right)=\max_{P_i:\sum_i P_i<P}\sum_{i=1}^{R_H}B\log_2\left(1+\frac{P_i\gamma_i}{P}\right) \tag{10-10}$$

其中$\gamma_i=\sigma_i^2P/\sigma^2$是满功率时第$i$个信道的信噪比。上式表明，高信噪比时，信道容量随信道的自由度线性增长；相反，低信噪比时，所有功率都会分配在信噪比最大（即σ_i^2最大）的那个信道上。式（10-10）与平衰落信道的式（4-9）或频率选择性衰落信道的式（4-23）类似。其最优解是MIMO信道的注水法功率分配：

$$\frac{P_i}{P}=\begin{cases}1/\gamma_0-1/\gamma_i & \gamma_i>\gamma_0\\ 0 & \gamma_i<\gamma_0\end{cases} \tag{10-11}$$

其中γ_0为某个门限值。由此得到容量为

$$C=\sum_{i:\gamma_i\geq\gamma_0}B\log_2\left(\frac{\gamma_i}{\gamma_0}\right) \tag{10-12}$$

例10.2：对于例10.1给出的MIMO信道，假设$\rho=P/\sigma_n^2=10\text{dB}$，$B=1\text{Hz}$，求信道容量和最优的功率分配。

解：由例10.1可知，信道奇异值分别为$\sigma_1=1.3333$、$\sigma_2=0.5129$、$\sigma_3=0.0965$。由$\gamma_i=10\sigma_i^2$可得$\gamma_1=17.7769$、$\gamma_2=2.6307$、$\gamma_3=0.0931$。假设所有并行信道都分配有功率，那么由功率约束条件可得

$$\sum_{i=1}^{3}\left(\frac{1}{\gamma_0}-\frac{1}{\gamma_i}\right)=1\implies\frac{3}{\gamma_0}=1+\sum_{i=1}^{3}\frac{1}{\gamma_i}=12.1749$$

解得$\gamma_0=0.2685$，这与$\gamma_3=0.0931<\gamma_0=0.2685$矛盾，因此第三个信道没有分配功率。再由功率约束条件得到

$$\sum_{i=1}^{2}\left(\frac{1}{\gamma_0}-\frac{1}{\gamma_i}\right)=1\implies\frac{2}{\gamma_0}=1+\sum_{i=1}^{2}\frac{1}{\gamma_i}=1.4364$$

解得$\gamma_0=1.392<\gamma_2$，这个γ_0即为正确的门限值。从而$P_i/P=1/1.392-1/\gamma_i$，因此$P_1/P=0.662$、$P_2/P=0.338$。容量为$C=\log_2(\gamma_1/\gamma_0)+\log_2(\gamma_2/\gamma_0)=4.59$。

对于有一个发送天线和多个接收天线的单入多出系统（single-input multiple-output，SIMO），或者有多个发送天线一个接收天线的多入单出系统（multiple-input single-output，MISO），也可以定义出收发都有理想信道信息时的容量。这些信道可以通过多天线获得分集增益和阵列增益，但没有复用增益。当发送端和接收端都已知信道信息时，其容量等于信号在发送端或接收端进行最大比合并后得到的SISO信道的容量，为$C=B\log_2(1+\rho\|\boldsymbol{h}\|^2)$，其中$\rho=P/\sigma^2$，信道矩阵$\boldsymbol{H}$变成了信道增益向量$\boldsymbol{h}$，最优权值向量

327 是$C=h^H/\|h\|$。

2. 发送端未知信道：平均功率分配

假设接收端已知信道信息，而发送端未知，那么发送端将无法在各天线上进行最优功率分配，或是优化天线之间的协方差特性。如果H分布符合ZMSW信道增益模型，其均值和方差对各个天线来说是对称的。因此我们自然会认为，应该把功率平均分配到每个发送天线上。这样，输入的协方差矩阵是酉阵乘上了一个系数：$R_x=(\rho/M_t)I_{M_t}$。文献[4]指出，在上述假设下，这样的输入协方差矩阵确实能使信道互信息最大化。此时，M_t发M_r收系统的互信息为

$$I(x;y)=B\log_2\det\left[\mathbf{I}_{M_r}+\frac{\rho}{M_t}\mathbf{H}\mathbf{H}^H\right]$$

可用奇异值分解将此式表示为

$$I(x;y)=\sum_{i=1}^{R_H}B\log_2\left(1+\frac{\gamma_i}{M_t}\right) \tag{10-13}$$

其中$\gamma_i=\sigma_i^2\rho=\sigma_i^2P/\sigma^2$。

式（10-13）所给出的MIMU信道的互信息与H，特别是奇异值$\{\sigma_i\}$的具体实现有关。对随机矩阵H平均后的平均互信息与H的概率分布有关[5],[11~12]。在衰落信道中，若发送端以此平均互信息为速率发送数据，可以保证接收端能正确接收，下一节对此进行讨论。对于静态信道，如果发送端不知道信道状态或者信道的平均互信息，那么它也无法确定该以什么样的速率发送方能保证数据的正确接收。此时最合适的容量定义应为中断容量。发送端以固定速率R来发送，中断率表示接收端不能正确接收的概率，也即信道H的互信息小于R的概率，其值为

$$P_{out}=p\left(H:B\log_2\det\left[I_{M_r}+\frac{\rho}{M_t}HH^H\right]<R\right) \tag{10-14}$$

这个概率取决于HH^H的特征值分布，这些特征值是H的奇异值的平方。人们对矩阵奇异值的分布问题已经进行了研究，对于MIMO信道中常见的情形，已经得到了分布的结果，328 参见参考文献[12]的2.1节。

当发送端和接收端的天线数目很多时，随机矩阵理论给出了关于H的奇异值分布的中心极限定理[13]。根据这一定理，所有信道实现都有相同的互信息。文献[14~17]用这些结论研究了独立衰落时的MIMO信道容量，文献[18~20]研究了相关衰落时的容量。作为极限分布的例子，假设接收端天线数固定为M_r，那么在ZMSW模型下，大数定律表明

$$\lim_{M_t\to\infty}\frac{1}{M_t}HH^H=I_{M_r} \tag{10-15}$$

将此式代入式（10-13）可得，当M_t趋于无限大时，互信息变为常数$C=M_r B\log_2(1+\rho)$。定义$M=\min(M_t, M_r)$，则随着M的增大，MIMO信道在未知CSIT的情况下容量将趋于$C=MB\log_2(1+\rho)$，随M线性增长。在天线数较少时也能观察到这种随M线性增长的规律[20]。同样的，当信噪比很大时，对于任意M_t和M_r，容量也随$M=\min(M_t, M_r)$线性增长[2]。由

于ZMSW MIMO信道的秩$R_H=M=\min(M_t, M_r)$，因此无CSIT时，高信噪比或者天线数很多时，信道容量随信道自由度线性增长。这些结论是MIMO技术颇具吸引力的主要原因：只要接收端能够正确估计信道信息，即使发送端不知道信道状态，ZMSW MIMO信道的容量也与发送端和接收端中的最小天线数成线性增长关系。因此MIMO信道在不需要增加信号功率或带宽的情况下就可以提供很高的数据率。但需注意，信噪比非常低时增加发送天线并无益处，容量只与接收天线数有关。这是因为信噪比非常低时，MIMO系统只是在集中能量，不能利用所有可用的自由度。此时，对于ZMSW MIMO信道，无论是把能量扩展到所有发送天线上，还是集中在一两根天线上都能达到容量[4]。随着信噪比的增加，限制因素不再是功率，而是信道的自由度。

尽管至少在天线数很多时，无CSIT并不影响容量的增长率，但它会增加解调的复杂度。因为发送端未知信道便无法将MIMO信道转化为互不干扰的SISO信道。前面已经提到，译码复杂度与所有发送天线上的独立符号数成指数关系，其中的独立符号数等于输入协方差矩阵的秩。

我们在这里的分析是在理想的CSIR和没有CSIT条件下假设了信道增益矩阵分布为ZMSW，即均值为零、协方差矩阵为单位阵。如果信道均值不为零，或者协方差矩阵不是单位阵的话，信道在空间上就存在差别，此时平均功率分配不是最优的方法，可以利用这种空间差别来采用最优的发送策略[21-23]。文献[21]、[24]的结论表明，当信道中存在主导的均值或协方差方向时，可用波束成形来达到信道容量，波束成形将在10.4节描述。因为波束成形实现起来非常简单，所以这种空间的差异性是一种非常有利的情形。 329

10.3.2 衰落信道

假设信道增益矩阵经历了平衰落，即增益h_{ij}随时间变化。类似于静态信道，信道容量取决于发送端和接收端是否已知信道矩阵。有理想的CSIR和CSIT时，发送端可以自适应于信道的衰落，此时的信道容量是最优功率分配下对所有信道矩阵实现的平均。有CSIR、无CSIT时，我们用遍历容量和中断容量来描述信道的容量。下面对此进行详细讨论。

1. 发送端已知信道：注水法

有CSIT和CSIR时，发送端将像静态信道那样根据每个衰落信道的实现优化发送策略。对每次实现的信道容量（10-8）进行平均就是信道的遍历容量。这个平均容量叫做信道遍历容量功率分配有两种可能。短期功率约束下，每个信道实现的发送功率等于平均功率$\overline{P}$，遍历容量为

$$\begin{aligned} C &= \boldsymbol{E}_H\left[\max_{\boldsymbol{R}_x:\mathrm{Tr}(\boldsymbol{R}_x)=\rho} B\log_2\det[\boldsymbol{I}_{M_r}+\boldsymbol{H}\boldsymbol{R}_x\boldsymbol{H}^H]\right] \\ &= \boldsymbol{E}_H\left[\max_{P_i:\Sigma_i P_i\leqslant \overline{P}} \sum_i B\log_2\left(1+\frac{P_i\gamma_i}{\overline{P}}\right)\right] \end{aligned} \tag{10-16}$$

其中$\gamma_i=\sigma_i^2\overline{P}/\sigma^2$。长期功率约束下，不同的信道实现$\boldsymbol{H}$对应不同的发送功率$P_H$，所有实现的平均功率满足$\boldsymbol{E}_H[P_H]\leqslant\overline{P}$，遍历容量为

$$
\begin{aligned}
C &= \max_{\rho_H: E_H[\rho_H]=\rho} E_H\left[\max_{\boldsymbol{R}_x:\mathrm{Tr}[\boldsymbol{R}_x]=\rho_H} B\log_2\det[\boldsymbol{I}_{M_r}+\boldsymbol{H}\boldsymbol{R}_x\boldsymbol{H}^H]\right] \\
&= \max_{P_H: E_H[P_H]\leqslant \bar{P}} E_H\left[\max_{P_i:\sum_i P_i \leqslant P_H} \sum_i B\log_2\left(1+\frac{P_i\gamma_i}{P_H}\right)\right]
\end{aligned}
\tag{10-17}
$$

其中 $\gamma_i=\sigma_i^2 P_H/\sigma^2$ 。短期功率约束是沿天线的空间注水，而长期功率约束是沿时间和空间的二维注水，类似于时变频率选择性衰落容量中的时频注水。式（10-16）和式（10-17）第二行中方括号内的项是$\boldsymbol{H}$的奇异值$\{\sigma_i\}$的函数，因此式中关于$\boldsymbol{H}$的数学期望取决于奇异值的分布。$\boldsymbol{H}$为ZMSW矩阵或其他类型矩阵时奇异值的分布可见文献[12]的2.1节。 330

2. 发送端未知信道：遍历容量和中断容量

考虑时变信道，假设接收端已知随机矩阵$\boldsymbol{H}$，但发送端未知，发送端假设$\boldsymbol{H}$服从ZMSW分布。此时有两种重要的容量定义，分别为遍历容量和中断容量。遍历容量是按照所有的信道实现平均后的最大可传输速率，它只取决于$\boldsymbol{H}$的分布。达到这个速率要求发送端在功率约束的条件下优化信道输入的协方差矩阵。这一问题的数学表示为

$$
C=\max_{\boldsymbol{R}_x:\mathrm{Tr}(\boldsymbol{R}_x)=\rho} E_H[B\log_2\det[\boldsymbol{I}_{M_r}+\boldsymbol{H}\boldsymbol{R}_x\boldsymbol{H}^H]] \tag{10-18}
$$

其中的数学期望是对信道矩阵$\boldsymbol{H}$的分布进行。对于ZMSW模型，$\boldsymbol{H}$是独立同分布的零均值循环对称的，其方差为1。

如同静态信道的结果，能使ZMSW信道的遍历容量最大化的输入协方差矩阵为 $\boldsymbol{R}_x=(\rho/M_t)\boldsymbol{I}_{M_t}$ ，就是将发送功率平分到各个天线，每个天线发送独立的符号。于是，遍历容量为

$$
C=E_H\left[B\log_2\det\left[\boldsymbol{I}_{M_r}+\frac{\rho}{M_t}\boldsymbol{H}\boldsymbol{H}^H\right]\right] \tag{10-19}
$$

其中的数学期望基于$\boldsymbol{H}$的奇异值分布，这一点和CSIT已知时一样。静态信道中的这个结论，容量随M=min（M_t, M_r)的增大而增大，也同样适用于遍历容量，因为它只是静态信道容量的简单平均。关于增长率的表达式可见文献[4]、[25]。当信道不是ZMSW时，容量取决于随机信道矩阵的奇异值分布，文献[11]研究了奇异值的分布及相应的遍历容量。

对于独立同分布复高斯信道增益的4×4MIMO系统，图10-4示出了接收端和发送端都已知CSI和仅接收端已知CSI时的遍历容量。如同SISO信道，这两条曲线差距很小，且此差距随信噪比的增大而减小。如果同图4-7所示的SISO衰落信道的容量相比，MIMO信道的容量增大了min (M_t, M_r)=4倍，与前面的结论吻合。 331

当发送端未知信道增益矩阵，而矩阵元素为复高斯但非独立同分布时，发送端可以利用信道的均值或协方差来提高容量。其基本思想是根据均值或协方差来分配功率，这种情形也称为均值或协方差反馈。假设接收端理想已知CSI，相关衰落的影响取决于发送端已知信道的情况：如果发送端知道信道的实现或者如果发送端既不知道信道的实现，也不知道信道的相关特性，那么天线之间的相关性将会使容量下降。反之，如果发送端知道相关特性，则当M_r=1时容量可以变大。详见文献[21]、[24]、[26]。

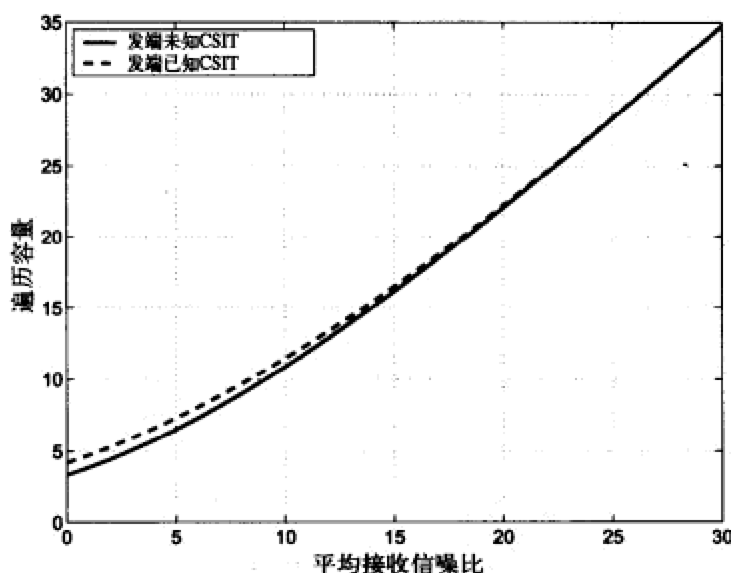


图10-4 4×4MIMO信道的遍历容量

假设信道是慢变的，信道矩阵H在相对较长的时间内保持不变，那么可类似10.3.1节中静态信道的情形给出中断容量的定义。正如静态信道一样，发送端并不知道信道的状态和相应的容量，但它必须要以一个固定的速率发送数据。无论发送端选择的速率R是多少，都会有发生中断的概率，即接收端不能正确接收的概率，中断率与静态信道的式（10-14）相同。给一个或多个天线不分配发送功率有时可以提高中断容量，尤其在中断率较高的时候[4]。这是因为中断容量取决于概率分布的拖尾，天线数较少时，所做的平均就越少，拖尾的扩展也相应变大。

332

对于独立同分布复高斯信道增益的4×4MIMO系统，图10-5示出了中断容量，中断率分别为1%和10%。可以看到，随着信噪比的增加，这两个中断率对应的中断容量的差距在增大。这一点可由图10-6所示的容量的分布曲线来解释。低信噪比时分布非常陡峭，因此中断率为1%和10%的容量很接近。高信噪比时，分布曲线的陡峭程度变缓，使得不同中断率的容量差别随之增大。

3. 发送端和接收端均未知CSI

当发送端和接收端均未知CSI时，容量不再与天线数成正比，有时增加天线反倒会降低容量。而且，信道容量越来越依赖于信道模型的选取，从而难以概括容量的增长。文献[27]指出，对于独立同分布的ZMSW分组衰落信道，增加的天线数如果超过分组的持续时间将不会增加容量，因此天线数超过一定数值时将不能提高数据速率。不过衰落有相关性时，增加发送天线的数目确实可以提高容量[28]。文献[29]将文献[27]的结论进行了扩展，针对高信噪比的独立同分布分组衰落，给出了显式的容量及达到容量的发送策略。文献[29]表明，对于独立同分布的衰落信道，容量随着信道的自由度线性增长。

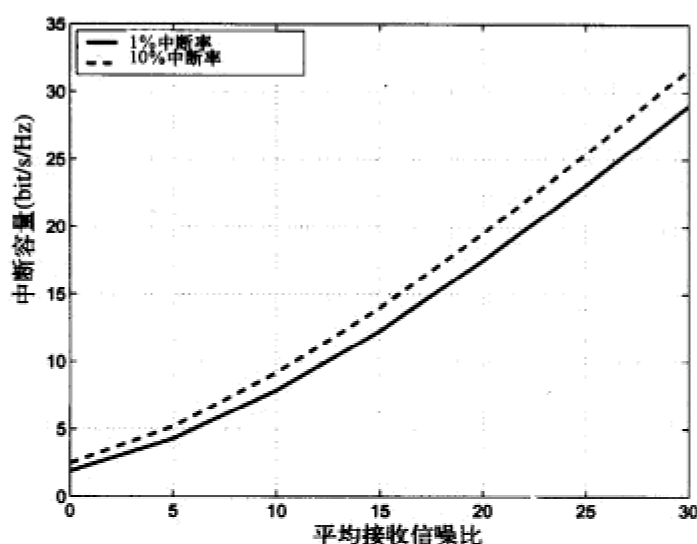


图10-5　4×4MIMO信道的中断容量

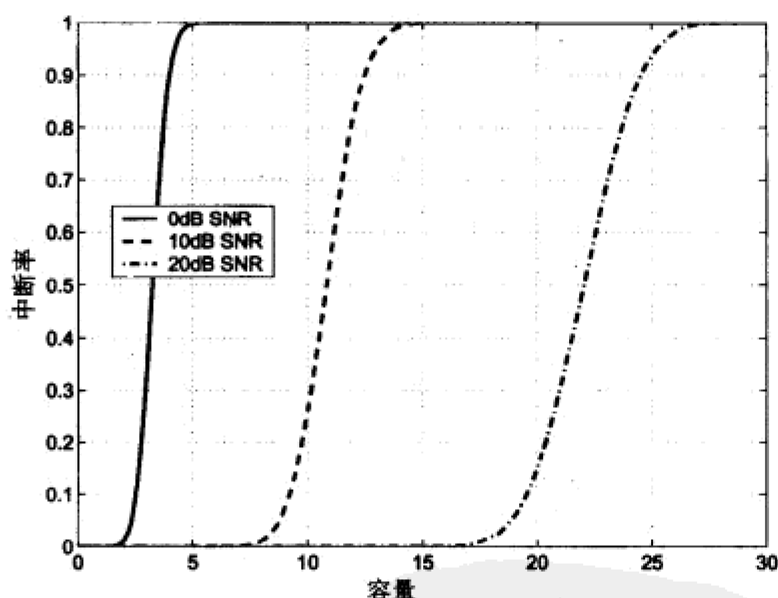


图10-6　4×4MIMO信道的中断率分布

对于接收端和发送端都未知CSI的情形，文献[30~32]研究了一般衰落模型下的容量问题。结果表明，低信噪比时容量受限于噪声，且随信道自由度线性增长；中等偏高信噪比时，容量受限于估计误差，且仍随信道自由度线性增长。信噪比很高时，自由度不再起作用，容量与信噪比成二重对数增长关系。换句话说，对于慢变衰落信道及收发都未知CSI的情形，高信噪比时增加天线并不能带来复用增益。

333

10.4 MIMO分集增益：波束成形

除了容量增益外，也可以用发送端和接收端的多天线来获得7.1节所定义的阵列和分集增益。此时，同一符号经不同的复数因子加权之后送到每个发送天线，输入协方差矩阵的秩为1。这一方法也称为MIMO波束成形[1]。波束成形策略是将10.2节所述的矩阵形式的预编码和接收成形变成了列向量形式：$\boldsymbol{V}=\boldsymbol{v}$，$\boldsymbol{U}=\boldsymbol{u}$，如图10-7所示。发送符号$x$经过$v_i$加权后在第$i$个天线上发送。接收端将第$i$个天线上的接收经过$u_i^*$加权。发送端和接收端的权值向量均归一化为$\|\boldsymbol{u}\|=\|\boldsymbol{v}\|=1$。最终形成的接收信号为

$$y=\boldsymbol{u}^H\boldsymbol{H}\boldsymbol{v}x+\boldsymbol{u}^H\boldsymbol{n} \tag{10-20}$$

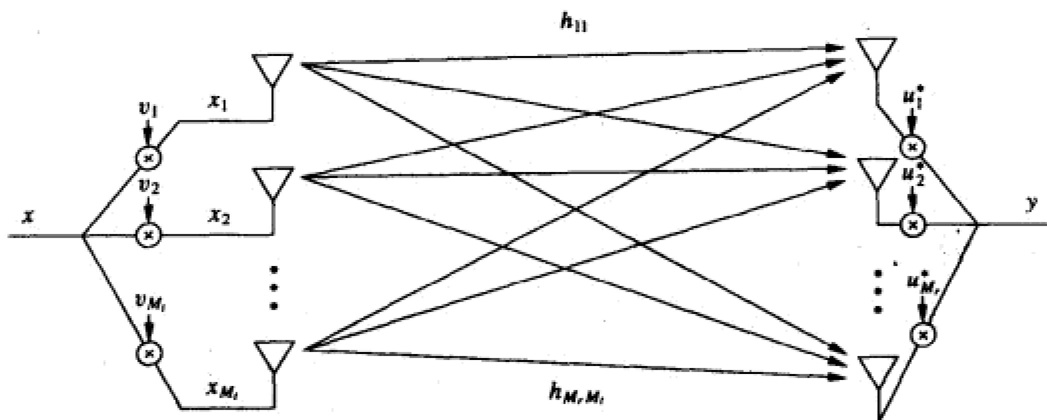


图10-7 波束成形的MIMO信道

若$\boldsymbol{n}=(n_1,\cdots,n_{M_r})$的元素为独立同分布，那么$\boldsymbol{u}^H\boldsymbol{n}$的统计特性与这些元素相同。

通过对不同路径的信号进行相干合并，波束成形可获得分集和阵列增益。相干合并已经要求接收端已知信道信息，因此性能增益就取决于发送端是否已知信道。当发送端已知信道矩阵$\boldsymbol{H}$时，能使接收信噪比最大的权值向量$\boldsymbol{u}$和$\boldsymbol{v}$分别是$\boldsymbol{H}$的主左奇异向量和主右奇异向量。即对于$\sigma_1=\sigma_{\max}$为$\boldsymbol{H}$最大的奇异值，$\boldsymbol{u}$和$\boldsymbol{v}$分别是$\boldsymbol{U}$和$\boldsymbol{V}$的第一列。相应的接收信噪比为[6, 33] $\gamma=\sigma_{\max}^2\rho$。相应的容量为$C=B\log_2(1+\sigma_{\max}^2\rho)$，等于功率增益为$\sigma_{\max}^2$的SISO信道容量。当$\boldsymbol{H}$是ZMSW矩阵时，波束成形可获得的阵列增益介于$\max(M_t, M_r)$和$M_t M_r$之间，分集增益为$M_t M_r$，详见[6]的5.4.4节。 334

当发送端未知信道时，若$M_t=2$，则可采用7.3.2节所述的Alamouti分集获得M_r的阵列增益和$2M_r$的分集增益，详见[6]的5.4.3节。若$M_t>2$，可用10.6.3节所述的空时分组码获得满分集增益。

虽然波束成形的容量不及发送预编码和接收成形，但它的译码复杂度是$|\chi|$数量级，

1. 波束成形这个术语也用在10.8节所述的智能天线中。在智能天线中，波束成形指将发送或接收天线对准某个给定的方向。

而非 $|\chi|^{R_H}$。一种更为简单的方法是收发的某一端采用最大比合并，另一端使用天线选择，详见文献[34]。

例10.3：已知MIMO信道的增益矩阵为

$$\boldsymbol{H}=\begin{bmatrix}0.7 & 0.9 & 0.8\\ 0.3 & 0.8 & 0.2\\ 0.1 & 0.3 & 0.9\end{bmatrix}$$

假设收发都已知信道，B=100kHz、ρ=10dB。求使用波束成形时的信道容量。

解：H的最大奇异值为 $\sigma_{\max}=\sqrt{\lambda_{\max}}$，其中$\lambda_{\max}$为如下矩阵的最大特征值

$$\boldsymbol{H}\boldsymbol{H}^H=\begin{bmatrix}1.94 & 1.09 & 1.06\\ 1.09 & 0.77 & 0.45\\ 1.06 & 0.45 & 0.91\end{bmatrix}$$

这个最大特征值为$\lambda_{\max}$=3.17，因此

$$C=B\log_2(1+\lambda_{\max}\rho)=10^5\log_2(1+31.7)=503\text{Kbps}$$

10.5　分集和复用的折中

前面介绍了两种利用多天线来提高无线系统性能的方法。一种方法是将MIMO信道分解为多个并行信道，在不同的信道上复用不同的数据流，这种增益称为复用增益。不过每个信道上的信噪比取决于信道矩阵的奇异值。容量分析中，信噪比差的信道体现为较低的数据率。对于实际的发送方式，若不采用强纠错能力的信道编码的话，这些信道的性能会很差。另一种方法是波束成形，它将信道增益进行相干合并，形成了一个分集增益很高的可靠信道。多天线不一定都用于复用或都用于分集，也可以用某些空时维作分集，其余的作复用。从而引出了一个MIMO系统设计的基本问题：多个天线是应该用于分集，复用，或是兼而有之？

335

关于MIMO系统中分集和复用的折中问题，或者更一般地说是数据率、错误率和复杂度的折中问题，已经有很多文献从理论角度和实际空时码的设计角度进行了广泛的研究[35-38]。这方面的工作主要考虑仅接收端已知CSI的分组衰落信道，因为收发都已知信道时的折中问题非常简单：先将天线子集分组来获得分集增益，然后用不同的分组获得复用增益。对于仅接收端已知CSI的分组衰落，当分组的长度趋于无限大时，可以通过天线间的对角编码，以合理的复杂度同时获得满分集增益和中断容量意义下的满复用增益[2, 39-40]。其中的一个例子是10.6.4节所述的D-BLAST。当分组长度有限时，不可能同时获得满分集和满复用增益，此时就需要在这两者之间进行折中。文献[38]对分组衰落信道在信噪比趋于无限大时的折中问题给出了一种简单的描述。其分析称一个系统具有复用增益r、分集增益d，若该系统每Hz的数据速率R(SNR)和错误概率P_e(SNR)满足

$$\lim_{\text{SNR}\to\infty}\frac{R(\text{SNR})}{\log_2\text{SNR}}=r \tag{10-21}$$

PDG

$$\lim_{\text{SNR}\to\infty}\frac{\log P_e(\text{SNR})}{\log \text{SNR}}=-d \tag{10-22}$$

式（10-22）中对数的底任意[1]。给定r，$d_{opt}(r)$表示最大可获得的分集增益。文献[38]证明，如果分组长度$T\geqslant M_t+M_r-1$，则有

$$d_{\text{opt}}(r)=(M_t-r)(M_r-r),0\leqslant r\leqslant \min(M_t,M_r) \tag{10-23}$$

如图10-8所示。此图表明：(i）如果将所有的发送和接收天线都用于分集，可以获得满分集增益M_tM_r；（ii）可以用一些天线牺牲一些分集增益来提高数据率。

336

也可以根据信道条件来调整分集和复用增益：在信道条件差的时候，多用一些天线作分集；在信道条件好的时候，用更多的天线作复用，见文献[41~43]。

例10.4：复用和分集参数r、d的定义如式（10-21）和（10-22）所示。假设高信噪比下近似有$d_{opt}(r)=(M_t-r)(M_r-r)$。设某个$M_t=M_r=8$的MIMO系统的信噪比为15dB，如果要求每赫兹的数据率为R=15bit/s，问系统可以提供的最大分集增益是多少？

解：已知SNR=15dB，欲R=15，需$r\log_2(10^{1.5})=15$，即r=3.01。因此3根天线用于复用，其余5根用于分集。从而最大的分集增益为$d_{opt}(r)=(M_t-r)(M_r-r)=(8-3)(8-3)=25$。

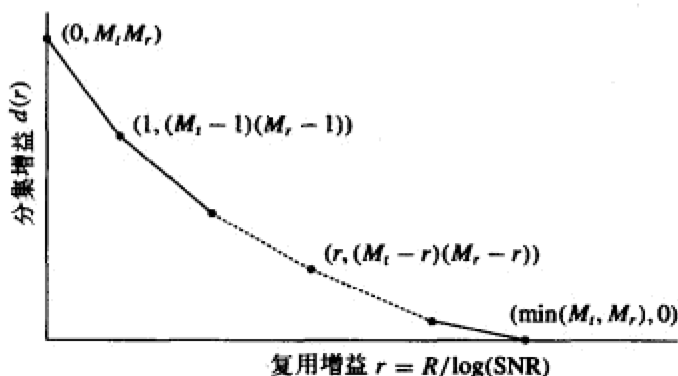


图10-8 高信噪比分组衰落时分集和复用的折中关系

10.6 空时调制与编码

传统SISO信道中，每个码元间隔内发送的是一个标量，但在MIMO信道中，输入输出关系为$\boldsymbol{y}=\boldsymbol{Hx}+\boldsymbol{n}$，每个码元间隔内发送的是一个向量。此外，MIMO中的信号设计可能包括多个天线和多个码元间隔，因此称为空时码（space-time code）。

包括本节在内的大部分空时码是为准静态信道（信道在T个码元周期内为常数）设计的，并假设发送端未知信道信息。在此模型下，信道的输入和输出是矩阵，其维数与空间（天线）和时间有关。令$\boldsymbol{X}=[x_1,\cdots,x_T]$表示$M_t\times T$的信道输入矩阵，其中的第$i$列$\boldsymbol{x}_i$是第$i$个传输时刻的信道输入向量。令$\boldsymbol{Y}=[y_1,\cdots,y_T]$表示$M_r\times T$的信道输出矩阵，其中的第$i$

1. 式（10-22）是两个同底的对数之比，其底可以约掉。

列y_i表示第i个传输时刻的信道输出向量。令$N=[n_1,\cdots,n_T]$表示$M_r\times T$的噪声矩阵，其中的第i列n_i表示第i个传输时刻的接收噪声向量。这样，T个码元周期内的输入输出关系为

$$Y=HX+N \tag{10-24}$$

10.6.1　最大似然检测及成对错误率

假设接收端已知信道矩阵H时，按照类似于第5章的标量情形和第8章的向量情形可以导出空时码的最大似然检测。对于给定的接收矩阵Y，最大似然发送矩阵 $\hat{X}$ 满足

337

$$\hat{X}=\arg\min_{X\in\mathcal{X}^{M_t\times T}}\|Y-HX\|_F^2=\arg\min_{X\in\mathcal{X}^{M_t\times T}}\sum_{i=1}^{T}\|y_i-Hx_i\|^2 \tag{10-25}$$

其中 $\|A\|_F$ 表示矩阵A的Frobenius范数[1]，上式是对所有可能的空时输入矩阵$\mathcal{X}^{M_t\times T}$求最小。将发送矩阵$X$错判为 $\hat{X}$ 的成对错误率 $p(\hat{X}\to X)$ 只取决于经过信道传输后这两个矩阵之间的距离以及噪声功率σ^2，即

$$p(\hat{X}\to X)=Q\left(\sqrt{\frac{\|H(X-\hat{X})\|_F^2}{2\sigma^2}}\right) \tag{10-26}$$

令 $D_X=X-\hat{X}$ 表示两个矩阵之差，由切尔诺夫界可得

$$p(\hat{X}\to X)\leqslant\exp\left[-\frac{\|HD_X\|_F^2}{4\sigma^2}\right] \tag{10-27}$$

令h_i表示H的第i行，$i=1,\cdots,M_r$，则

$$\|HD_X\|_F^2=\sum_{i=1}^{M_r}h_iD_XD_X^Hh_i^H \tag{10-28}$$

令$\mathcal{H}=\mathrm{vec}(H^T)^T$，其中vec($A$)表示将矩阵$A$的列上下排列所形成的列向量[2]，则$\mathcal{H}^T$是一个长为$M_tM_r$的列向量。同时定义$\mathcal{D}_X=I_{M_r}\otimes D_X$，其中$\otimes$表示Kronecker积，于是有

$$\|HD_X\|_F^2=\|\mathcal{H}\mathcal{D}_X\|_F^2 \tag{10-29}$$

代入式（10-27），并对所有可能的信道实现作数学期望得到

$$p(X\to\hat{X})\leqslant\left(\det\left[I_{M_tM_r}+\frac{1}{4\sigma^2}E[\mathcal{D}_X^H\mathcal{H}^H\mathcal{H}\mathcal{D}_X]\right]\right)^{-1} \tag{10-30}$$

假设信道矩阵H服从空间白随机模型，则其元素为独立同分布的零均、单位方差的复高斯随机变量。这样式（10-30）的数学期望成为

1. Frobenius范数是矩阵元素的模平方之和的平方根。

2. 对于$M\times N$的矩阵$A=[a_1,\cdots,a_N]$，其中a_i是长为M的列向量，$\mathrm{vec}(A)=[a_1^T,\cdots,a_N^T]^T$是长为$MN$的列向量。

$$p(X \to \hat{X}) \leqslant \left(\frac{1}{\det[I_{M_t} + 0.25\rho\Delta]}\right)^{M_r} \tag{10-31}$$

其中 $\Delta = (1/P)D_x D_x^H$ 。此式可以简化为

$$p(X \to \hat{X}) \leqslant \prod_{k=1}^{R_\Delta}\left(\frac{1}{1+0.25\gamma\lambda_k(\Delta)}\right)^{M_r} \tag{10-32}$$

338

其中$\gamma=P/\sigma^2=\rho$是每输入符号的信噪比，$\lambda_k(\Delta)$表示Δ的第k个非零特征值，$k=1,\cdots,R_\Delta$，R_Δ为Δ的秩。高信噪比（$\gamma>>1$）时，此式简化为

$$p(X \to \hat{X}) \leqslant \left(\prod_{k=1}^{R_\Delta}\lambda_k(\Delta)\right)^{-M_r}(0.25\gamma)^{-R_\Delta M_r} \tag{10-33}$$

从这个公式可以得到下节要讨论的空时码设计准则。

10.6.2 秩和行列式准则

式（10-33）表明成对错误率随γ^{-d}减小，其中$d=R_\Delta M_r$，因此$R_\Delta M_r$就是空时码的分集增益。M_t个发天线和M_r个收天线最大可能的分集增益为$M_t M_r$，因此，空时码要想获得最大的分集增益，必须把任意两码字的相差矩阵Δ设计为满秩M_t。这样的设计准则称为秩准则。

与式（10-33）中的对错误率相关的编码增益取决于$\left(\prod_{k=1}^{R_\Delta}\lambda_k(\Delta)\right)^{M_r}$这一项。因此，为了提高空时码的编码增益，必须要使所有输入矩阵对X和$\hat{X}$的差Δ中，最小的那个行列式最大化。这样的设计准则称为行列式准则。

秩准则和行列式准则最早在文献[36]、[44~45]中提出。和传统的二进制编码不同，这些准则是基于不同的发送矩阵之间的成对错误率的，往往需要通过计算机搜索来得到好码[46~47]。文献[48]提出了一种通用的二元秩准则，它给出了一种较好的构造空时码的方法。

10.6.3 空时格码和空时分组码

秩准则和行列式准则主要用于设计空时格码（space-time trellis code，STTC），它是卷积格码在MIMO系统中的扩展[6, 45]。它用格图来描述码，并用维特比译码进行最大似然序列估计。STTC有优良的分集增益和编码增益，但其译码复杂度随着分集程度和发送速率的增长而成指数增长[49]。空时分组码（space-time block code，STBC）是另外一种可选的空时码，也能获得显著的分集增益，而译码复杂度只是线性增长。对空时分组码的研究源于7.3.2所介绍的Alamouti码，对于两天线系统，它能获得满分集增益，且接收端只需做线性处理。文献[50]将Alamouti的编码扩展为一般的STBC，它对于任意数目的发送天线都可以获得满分集增益。不过这些空时分组码虽然获得了满增益，但没有编码增益，所以性能不如空时格码。可对空时格码或空时分组码以串行或并行的方式级联一个外码，以构成Turbo码来获得额外的编码增益[51~53]。文献[50]中的空时分组码的码字矩阵的各维是正交的，所以译码复杂度是线性的。文献[54]利用相同的思路设计出

了酉空时调制，它适合于收发都不知道信道信息的分组衰落信道。关于空时码更全面的介绍可参考文献[6]、[49]、[55~56]及其中的参考文献。

339

10.6.4　空分复用和BLAST结构

空分复用的基本思路是利用空间维数，在每个码元周期内发送M_t个独立的符号。为了获得满分集阶数，必须要将已编码的数据流在所有M_t个天线上发送。其实现方式可以是图10-9所示的串行编码。数据流先在信道分组时间长度T内进行编码形成码字$[x_1,\cdots,x_T]$，经交织后映射到星座点上，然后分路到不同的天线上。第一个码元周期内，前M_t个符号在M_t个天线上发送，下面M_t个符号在下一码元周期内发送，直到发完整个码字。i时刻在第k个天线上发送的符号记为$x_k[i]$。如果码字足够长，可以在所有M_t个天线上发送，M_r个天线上接收，这样就能获得满分集增益。不过，获得此增益需要的码字长度T为M_tM_r，而译码复杂度与这个长度成指数增长关系，这一点妨碍了串行编码的实用性。

340

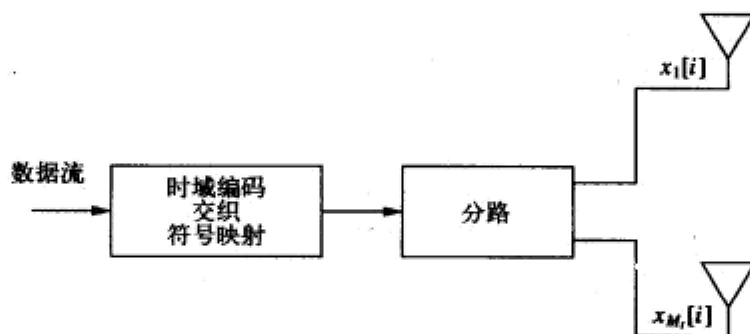


图10-9　串行编码的空分复用

更为简单的空分复用方法是并行编码，它是贝尔实验室所开创的贝尔实验室分层空时（Bell Labs Layered Space Time，BLAST）结构中的一种[2]，如图10-10所示。输入数据流被分为M_t个独立的子数据流，每个子数据流经过分组长度为T的单入单出编码、交织，然后映射到星座点上，并在相应的天线上发送。第k个单入单出编码器所产生的码字$\{x_k[i], i=1,\cdots, T\}$在第$k$个天线上按时间顺序发送。这个过程可以看成是将串行的数据流编码为垂直向量，故又称为**垂直编码**（vertical encoding）或V-BLAST[57]。因为每个编码由一个天线发送、M_r个天线接收，因此垂直编码的最大分集增益为M_r。它的编码复杂度与天线数目成线性增长关系，但最佳译码对各个发送天线上的码字进行联合检测，因为每个天线所发送的符号会被所有天线收到。文献[58]指出，采用图10-11所示的符号干扰抵消技术可以大大降低译码复杂度。由于各发送天线上符号是同步的，所以干扰抵消是这样进行的：先将M_t个发送符号按接收信噪比排序，对信噪比最高的符号进行估计，估计时将其他符号视为噪声。然后减去这个已估计的符号，再检测下一个信噪比最高的符号。这个过程持续到检测出所有M_t个符号为止。然后再对各子数据流分别进行译码，从而使译码复杂度与发送天线数成线性关系。这个结构实际上可以不用编码，

在文献[57]的报道中，未编码的V-BLAST在可接受的误比特率下可达到20bit/s/Hz~40bit/s/Hz的数据率。

341

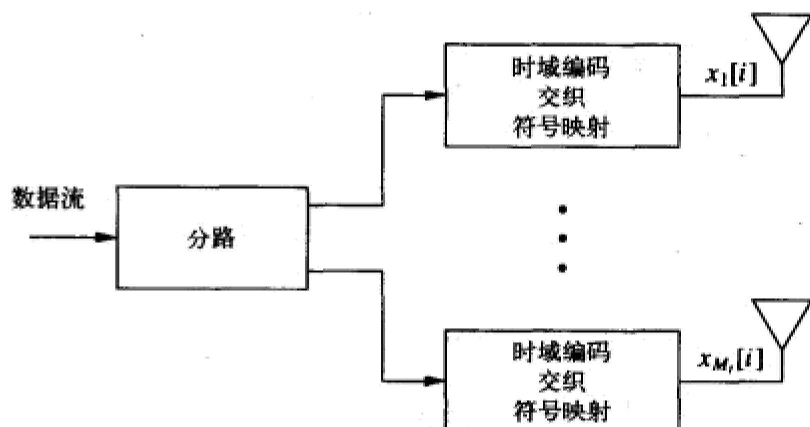


图10-10 并行编码的空分复用：V-BLAST

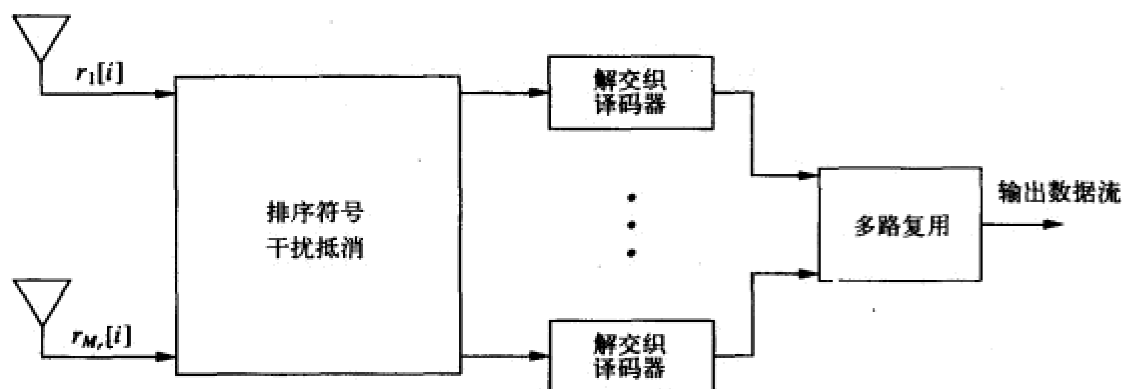


图10-11 线性复杂度的V-BLAST接收机

对角编码（diagonal encoding）或D-BLAST创造性地结合了并行编码的简单性和串行编码的分集增益[2]，如图10-12所示。D-BLAST先对数据流进行并行编码，但编码码字不是在一个天线上发送，而是以旋转的方式在所有天线上发送，从而使整个码字能在所有M_t个天线上发送。数据流的旋转如图10-13所示。假设第i个编码器产生的码字为$x_i=[x_i[1],\cdots,x_i[T]]$，那么$i$时刻在第一个天线上发送$x_i[1]$，$i$+1时刻在第二个天线上发送$x_i[2]$，依此类推。如果编码的分组长度$T$超过$M_t$，则再从第一个天线开始循环。最终使码字扩展到所有空间维。如果时域编码可达到容量（高斯码字，码长T无限大），那么基于D-BLAST的发送方案可以获得满分集增益M_tM_r，见文献[6]的6.3.5节。此外，如果对角上浪费的空时维可以忽略的话，D-BLAST可以获得最大的中断容量，见文献[6]的12.4.1节。因为接收端对每一对角码字单独译码，所以D-BLAST的接收复杂度也与发送天线数成线性关系。D-BLAST的缺点是图10-13所示的空时维浪费，如果帧的大小选择不当的话，这种浪费可能会很大。

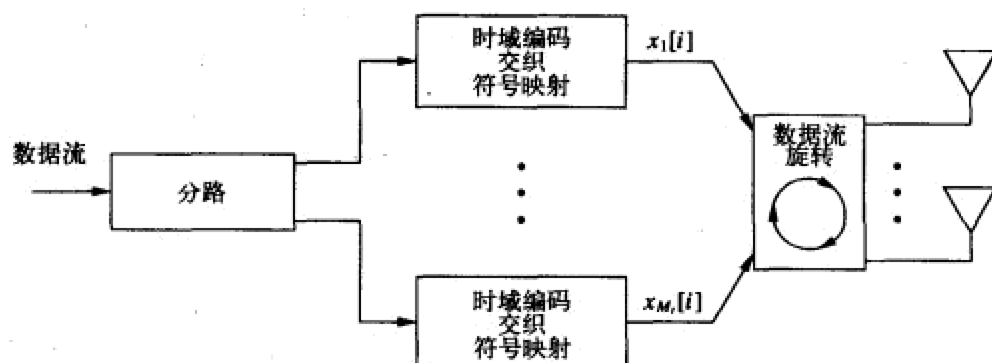


图10-12 带有数据流旋转的对角编码

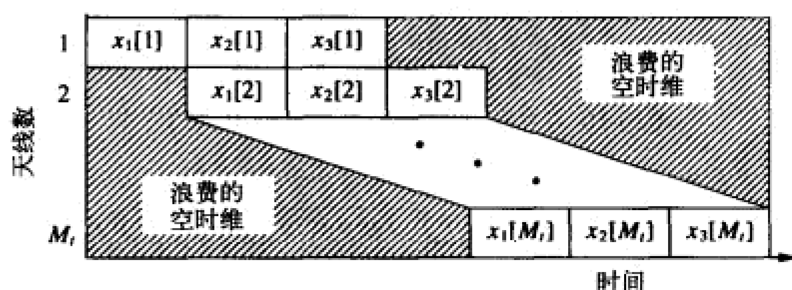


图10-13 数据流旋转

10.7 频率选择性MIMO信道

当MIMO信道的带宽大于多径时延扩展的倒数时，信道中将会产生ISI，这一点类似于SISO信道。有两种方法可以解决MIMO信道中的ISI。一种方法是采用信道均衡，不过MIMO信道中的均衡器必须要同时均衡时域和频域，因此非常复杂。此外，当均衡器和空时码结合使用时，码字的非线性和非因果性会使均衡器的设计更为复杂。在某些情况下，利用码字的结构特性可以将MIMO均衡问题转化为SISO均衡问题，从而能利用成熟的SISO均衡器[59-61]。

342

另一种解决频率选择性衰落问题的方法是采用多载波调制或正交频分复用(orthogonal frequency division multipiexing, OFDM)。第12章将讨论SISO信道中的OFDM技术，其主要思想是将宽带信道转化为一组平衰落的窄带子信道。将OFDM应用在MIMO信道中将产生一组窄带的MIMO信道，前面讨论的单个MIMO信道中的空时调制和编码可以直接用在这组并行信道中。频率选择性MIMO衰落信道在空间、时域、频域上均可获得分集，理想设计应当充分利用这三维。

10.8 智能天线

前面我们看到，通过空时信号处理，利用收发两端的多个天线可以提高数据率，也

可以提供分集增益。此外，发送端或者接收端的天线阵列也可通过扇区化或相控阵技术提供方向性天线增益。天线的方向性可以增加覆盖范围、减轻时延扩展（ISI）和平衰落、抑制用户间的干扰。实际中的干扰通常是从不同的方向到达的，因此定向天线可以利用这种方向上的差别来减少或消除来特定方向上的干扰，从而提高系统的容量。反射路径上的分量也是从不同方向到达接收端的，因此也可以得到衰减，从而减小ISI和平衰落。多天线系统可用来获得方向性增益、或者分集、或者复用增益，需要权衡运用。到底应该通过复用来提高数据率，还是通过分集来增加抗衰落性，亦或通过方向性来减小ISI和干扰，是一个复杂的问题，需要考虑系统的整体设计。

最常用的定向天线是扇区或相控（定向）天线阵列，这些天线以及全向天线的方向图如图10-14所示。扇区天线对某个到达角范围内的信号有很高的增益。扇区天线常用在蜂窝系统的基站中以减小干扰：假设不同的扇区分配有相同的频率或时隙，理想扇区化后，只有本扇区的用户间之间才会产生干扰，从而使系统的平均干扰减少为扇区数分之一。例如对于图10-14所示的120°波束宽度的扇区，基站将360°方向划分为3个扇区，每个扇区用一个120°的扇区天线覆盖，这样每一扇区的干扰减少为全向天线时的1/3。蜂窝系统用扇区化降低干扰的同时付出的代价是需要扇区间的切换。 343

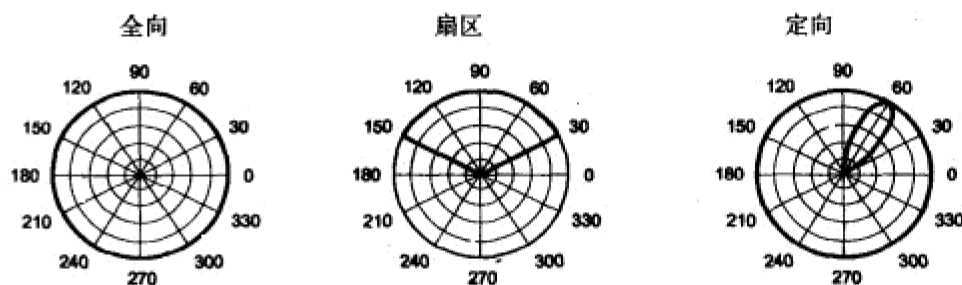


图10-14 全向天线、扇区天线和定向天线的方向图

有向天线一般用天线阵列结合相控阵技术来提供方向性增益，这可以由充分多的天线阵元来紧紧控制。相控阵技术通过调整每一天线阵元的相位来控制天线的方向，它可以控制天线的波束方向和零陷方向，波束方向是天线增益最大的方向，零陷方向是增益最小的方向。阵元个数越多，方向的控制就越准确。对于有N个天线的阵列，可以形成N个零陷，从而能显著降低来自N个独立方向的干扰功率。如果系统中有$N_I<N$个干扰源，则可用相控阵列中的N_I个天线来消除这N_I个干扰，用其余$N-N_I$个天线进行分集。注意，定向天线必须要知道有用信号和干扰信号的角度，以便在特定方向上产生高增益或低增益。对于高速移动的系统，跟踪用户的方向非常困难，所以蜂窝系统的基站采用了扇区天线，而不是智能天线。

天线阵列处理的复杂性以及安装天线阵列所需要的空间使得近期内智能天线不大可能用在体积小、重量轻、低功耗的手持终端设备上。不过，许多系统中的基站和接入点已经采用了天线阵列技术。有关智能天线及其在无线通信中的应用请参考文献[62]。

习题

10.1 MIMO信道分析中经常要用到一些矩阵方面的等式，请证明下面的矩阵等式。

(a) 给定$M\times N$的矩阵$\boldsymbol{A}$，证明$\boldsymbol{AA}^H$为复共扼。$\boldsymbol{AA}^H$的特征值分解有何特性？

(b) 证明$\boldsymbol{AA}^H$半正定。

(c) 证明$\boldsymbol{I}_M+\boldsymbol{AA}^H$为复共扼正定。

(d) 证明$\det[\boldsymbol{I}_M+\boldsymbol{AA}^H]=\det[\boldsymbol{I}_N+\boldsymbol{A}^H\boldsymbol{A}]$。

10.2 求下面矩阵的奇异值分解：

$$\boldsymbol{H}=\begin{bmatrix}0.7 & 0.6 & 0.2 & 0.4\\ 0.1 & 0.5 & 0.9 & 0.2\\ 0.3 & 0.6 & 0.9 & 0.1\end{bmatrix}$$

10.3 给出一个3×3的信道矩阵$\boldsymbol{H}$，它有两个非零奇异值。

10.4 考虑下面所列的4×4MIMO信道。每个信道的最大复用增益是多少？即信道可支 344 持多少个独立数据流的可靠传输？

$$\boldsymbol{H}_1=\begin{bmatrix}1 & 1 & -1 & 1\\ 1 & 1 & -1 & -1\\ 1 & 1 & 1 & 1\\ 1 & 1 & 1 & -1\end{bmatrix}\qquad \boldsymbol{H}_2=\begin{bmatrix}1 & 1 & 1 & -1\\ 1 & 1 & -1 & 1\\ 1 & -1 & 1 & 1\\ 1 & -1 & -1 & -1\end{bmatrix}$$

10.5 仅接收端已知CSI的静态MIMO信道的容量为 $C=\sum_{i=1}^{R_H}\log_2(1+\sigma_i^2\rho/M_t)$ 。证明：如果奇异值之和受限，那么此式在所有R_H个奇异值相等时最大。

10.6 假设MIMO系统的信道矩阵为：

$$\boldsymbol{H}=\begin{bmatrix}0.1 & 0.3 & 0.4\\ 0.3 & 0.2 & 0.2\\ 0.1 & 0.3 & 0.7\end{bmatrix}=\begin{bmatrix}-0.5196 & -0.0252 & -0.8541\\ -0.3460 & -0.9077 & 0.2372\\ -0.7812 & -0.4188 & -0.4629\end{bmatrix}\begin{bmatrix}0.9719 & 0 & 0\\ 0 & 0.2619 & 0\\ 0 & 0 & 0.0825\end{bmatrix}$$
$$\times\begin{bmatrix}-0.2406 & -0.4727 & -0.8477\\ -0.8894 & -0.2423 & 0.3876\\ 0.3886 & -0.8472 & 0.3621\end{bmatrix}^H$$

注意$\boldsymbol{H}$已表示为奇异值分解的形式：$\boldsymbol{H}=\boldsymbol{U}\Sigma\boldsymbol{V}^H$。

(a) 验证$\boldsymbol{H}=\boldsymbol{U}\Sigma\boldsymbol{V}^H$是否成立。你会发现矩阵$\boldsymbol{U}$、$\Sigma$和$\boldsymbol{V}^H$并非完全精确，$\boldsymbol{U}\Sigma\boldsymbol{V}^H$只是近似等于$\boldsymbol{H}$。这个结果表明了奇异值分解，特别是$\Sigma$，对信道矩阵$\boldsymbol{H}$的估计误差的敏感性。

(b) 根据奇异值分解$\boldsymbol{H}=\boldsymbol{U}\Sigma\boldsymbol{V}^H$，给出与此MIMO系统等效的三个独立信道，并求能将原系统变换为这些独立信道所需要的发送预编码矩阵及接收成形矩阵。

(c) 假设P/σ^2=20dB，系统带宽B=100kHz，给出对（b）中得到的三个独立信道的最优的功率分配P_i (i=1,2,3)，以及相应的容量。

(d) 发送端未知信道时功率将在三个天线间平分，求此时的容量，并与（c）中结果比较。

10.7 利用奇异值分解的性质证明：当收发都已知信道时，MIMO信道容量的一般表达式

$$C=\max_{\boldsymbol{R}_x:Tr(\boldsymbol{R}_x)=\rho}B\log_2\det[\boldsymbol{I}_{M_r}+\boldsymbol{HR}_x\boldsymbol{H}^H]$$

可简化为

$$C = \max_{\rho_i : \sum_i \rho_i \le \rho} \sum_i B \log_2(1 + \sigma_i^2 \rho_i)$$

其中$\{\sigma_i\}$是奇异值，ρ是信噪比。

10.8　对于下面给出的4×4MIMO信道，假设收发两端都已知信道信息，信噪比ρ=10dB，带宽B=10MHZ，求信道容量。

$$\boldsymbol{H}_1 = \begin{bmatrix} 1 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & -1 \end{bmatrix} \qquad \boldsymbol{H}_2 = \begin{bmatrix} 1 & 1 & 1 & -1 \\ 1 & 1 & -1 & 1 \\ 1 & -1 & 1 & 1 \\ 1 & -1 & -1 & -1 \end{bmatrix}$$

10.9　某ZMCSCG MIMO系统中发送和接收天线数为$M_t=M_r=M$。利用大数定律证明

$$\lim_{M\to\infty} \frac{1}{M}\boldsymbol{H}\boldsymbol{H}^H = \boldsymbol{I}_M$$

其中$\boldsymbol{H}$是信道矩阵。再由此证明

$$\lim_{M\to\infty} B \log_2 \det\left[\boldsymbol{I}_M + \frac{\rho}{M}\boldsymbol{H}\boldsymbol{H}^H\right] = MB\log_2(1+\rho)$$

345

10.10　按照信噪比范围及$0\le\rho\le30$dB及B=1MHz，画出下列情形时ZMCSCG MIMO信道的遍历容量：(a) $M_t=M_r=1$；(b) $M_t=2, M_r=1$；(c) $M_t=M_r=2$；(d) $M_t=2, M_r=3$；(e) $M_t=M_r=3$。并验证在高信噪比下，容量随$M=\min(M_t, M_r)$线性增长。

10.11　按照信噪比范围$0\le\rho\le30$dB、B=1MHz、中断率P_{out}=0.01，画出下列情形时ZMCSCG MIMO信道的中断容量：(a) $M_t=M_r=1$；(b) $M_t=2$，$M_r=1$；(c) $M_t=M_r=2$；(d) $M_t=2$，$M_r=3$；(e) $M_t=M_r=3$。并验证在高信噪比下，容量随$M=\min(M_t, M_r)$线性增长。

10.12　证明：若噪声向量$\boldsymbol{n}=(n_1, \cdots n_{M_r})$的元素为独立同分布，则对于$\|\boldsymbol{u}\|=1$，$\boldsymbol{u}^H\boldsymbol{n}$的统计特性和$\boldsymbol{n}$的元素相同。

10.13　某MIMO系统中收发两端都已知信道增益矩阵$\boldsymbol{H}$。证明：若发送和接收天线都用于分集，那么收发都采用最优权值时的信噪比为$\gamma=\lambda_{max}\rho$，其中λ_{max}为$\boldsymbol{H}\boldsymbol{H}^H$的最大特征值。

10.14　假设信道矩阵为

$$\boldsymbol{H} = \begin{bmatrix} 0.1 & 0.5 & 0.9 \\ 0.3 & 0.2 & 0.6 \\ 0.1 & 0.3 & 0.7 \end{bmatrix}$$

ρ=10dB。求发送端权值相等，接收端用最优权值时，波束成形的输出信噪比。并与收发都用最优权值时的信噪比进行比较。

10.15　假设某8×4 MIMO系统所采用的编码满足折中关系$d(r)=(M_t-r)(M_r-r)$。

(a) 若ρ=10dB，要求$P_e=\rho^{-d}\le10^{-3}$，求信道的最大复用率。

(b) 对于 (a) 中得到的r，对应的P_e是多少？

10.16　设SIMO信道的增益向量为$\boldsymbol{h}$=[0.1　0.4　0.75　0.9]，ρ=10dB，B=10MHz，求接

收端使用最优权值时的信道的容量。

10.17　某2×2 MIMO系统的信道增益矩阵$\boldsymbol{H}$为

$$\boldsymbol{H}=\begin{bmatrix}0.3 & 0.5\\0.7 & 0.2\end{bmatrix}$$

假设收发两端都已知$\boldsymbol{H}$，总发送功率为P=10mW，每个接收天线上的加性白高斯噪声的单边功率谱密度为$N_0=10^{-9}$W/Hz，信道带宽为B=100kHZ。

（a）求$\boldsymbol{H}$的奇异值分解。

（b）求信道容量。

（c）假设用发送预编码和接收成形将此信道变换为两个并行独立信道，总发送功率P在两信道间最优分配。每一信道采用MQAM调制，星座大小不限，BER界为$P_b\leqslant 0.2e^{-1.5\gamma/(M-1)}$。目标BER为$10^{-3}$。求这组并行信道上可传输的最大数据速率。

346

（d）假设发送端和接收端的天线都用于分集，且都使用最优权值以最大化合并输出的信噪比。求合并输出的信噪比以及BPSK调制的误比特率。假设$D=1/T_b$，请将此BPSK的数据速率和误比特率与（c）中的结果进行比较。

（e）就分集和复用的问题，讨论（c）和（d）两个系统之间如何权衡。

10.18　假设$M\times M$ MIMO信道的增益服从ZMCSCG分布。

（a）若发送端和接收端都已知CSI，B=1MHz，请按信噪比范围$0\leqslant\rho\leqslant 20$dB画出$M$=1和$M$=4时的遍历容量。

（b）若只有接收端有CSI时，重复（a）。

10.19　假设4×4 MIMO信道的增益服从ZMCSCG分布。求中断率为10%、ρ=10dB、B=1MHz时的中断容量。

10.20　若$M\times M$ MIMO信道的ρ=10dB，B=1MHz，发送端无信道信息。请画出M=4、6、8时容量的累积分布函数。M增大时曲线有何变化？这对实际系统设计意味着什么？

参考文献

[1] J. Winters, "On the capacity of radio communication systems with diversity in a Rayleigh fading environment," *IEEE J. Sel. Areas Commun.*, pp. 871–8, June 1987.

[2] G. J. Foschini, "Layered space-time architecture for wireless communication in fading environments when using multi-element antennas," *Bell System Tech. J.*, pp. 41–59, Autumn 1996.

[3] G. J. Foschini and M. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Pers. Commun.*, pp. 311–35, March 1998.

[4] E. Telatar, "Capacity of multi-antenna Gaussian channels," AT&T Bell Labs Internal Tech. Memo, June 1995.

[5] E. Telatar, "Capacity of multi-antenna Gaussian channels," *Euro. Trans. Telecommun.*, pp. 585–96, November 1999.

[6] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*, Cambridge University Press, 2003.

[7] L. H. Brandenburg and A. D. Wyner, "Capacity of the Gaussian channel with memory: The multivariate case," *Bell System Tech. J.*, pp. 745–78, May/June 1974.

[8] J. Salz and A. D. Wyner, "On data transmission over cross coupled multi-input, multi-output lin-

ear channels with applications to mobile radio," AT&T Bell Labs Internal Tech. Memo, 1990.
[9] B. Tsybakov, "The capacity of a memoryless Gaussian vector channel," *Prob. Inform. Trans.*, 1(1), pp. 18–29, 1965.
[10] J. L. Holsinger, "Digital communication over fixed time-continuous channels with memory, with special application to telephone channels," MIT Res. Lab Elec. Tech. Rep. 430, 1964. 347
[11] H. Shin and J. H. Lee, "Capacity of multiple-antenna fading channels: Spatial fading correlation, double scattering, and keyhole," *IEEE Trans. Inform. Theory*, pp. 2636–47, October 2003.
[12] A. M. Tulino and S. Verdú, "Random matrix theory and wireless communications," *Found. Trends Commun. Inform. Theory*, 1(1), pp. 1–182, 2004.
[13] V. L. Girko, "A refinement of the central limit theorem for random determinants," *Theory Probab. Appl.*, 42(1), pp. 121–9, 1998.
[14] A. Grant, "Rayleigh fading multiple-antenna channels," *J. Appl. Signal Proc.*, Special Issue on Space-Time Coding (Part I), pp. 316–29, March 2002.
[15] P. J. Smith and M. Shafi, "On a Gaussian approximation to the capacity of wireless MIMO systems," *Proc. IEEE Internat. Conf. Commun.*, pp. 406–10, April 2002.
[16] S. Verdú and S. Shamai (Shitz), "Spectral efficiency of CDMA with random spreading," *IEEE Trans. Inform. Theory*, pp. 622–40, March 1999.
[17] Z. Wang and G. B. Giannakis, "Outage mutual information of space-time MIMO channels," *Proc. Allerton Conf. Commun., Control, Comput.*, pp. 885–94, October 2002.
[18] C.-N. Chuah, D. N. C. Tse, J. M. Kahn, and R. A. Valenzuela, "Capacity scaling in MIMO wireless systems under correlated fading," *IEEE Trans. Inform. Theory*, pp. 637–50, March 2002.
[19] A. Lozano, A. M. Tulino, and S. Verdú, "Multiple-antenna capacity in the low-power regime," *IEEE Trans. Inform. Theory*, pp. 2527–44, October 2003.
[20] A. L. Moustakas, S. H. Simon, and A. M. Sengupta, "MIMO capacity through correlated channels in the presence of correlated interferers and noise: A (not so) large N analysis," *IEEE Trans. Inform. Theory*, pp. 2545–61, October 2003.
[21] S. A. Jafar and A. J. Goldsmith, "Transmitter optimization and optimality of beamforming for multiple antenna systems," *IEEE Trans. Wireless Commun.*, pp. 1165–75, July 2004.
[22] A. Narula, M. Lopez, M. Trott, and G. Wornell, "Efficient use of side information in multiple-antenna data transmission over fading channels," *IEEE J. Sel. Areas Commun.*, pp. 1423–36, October 1998.
[23] E. Visotsky and U. Madhow, "Space-time transmit precoding with imperfect feedback," *Proc. IEEE Internat. Sympos. Inform. Theory*, pp. 357–66, June 2000.
[24] E. Jorswieck and H. Boche, "Channel capacity and capacity-range of beamforming in MIMO wireless systems under correlated fading with covariance feedback," *IEEE Trans. Wireless Commun.*, pp. 1543–53, September 2004.
[25] B. Hochwald and V. Tarokh, "Multiple-antenna channel hardening and its implications for rate feedback and scheduling," *IEEE Trans. Inform. Theory*, pp. 1893–1909, September 2004.
[26] A. J. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE J. Sel. Areas Commun.*, pp. 684–701, June 2003.
[27] T. Marzetta and B. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Trans. Inform. Theory*, pp. 139–57, January 1999.
[28] S. A. Jafar and A. J. Goldsmith, "Multiple-antenna capacity in correlated Rayleigh fading with channel covariance information," *IEEE Trans. Wireless Commun.*, pp. 990–7, May 2005.
[29] L. Zheng and D. N. Tse, "Communication on the Grassmann manifold: A geometric approach to the non-coherent multi-antenna channel," *IEEE Trans. Inform. Theory*, pp. 359–83, February 2002.
[30] R. Etkin and D. Tse, "Degrees of freedom in underspread MIMO fading channels," *Proc. IEEE*

Internat. Sympos. Inform. Theory, p. 323, July 2003.

[31] A. Lapidoth and S. Moser, "On the fading number of multi-antenna systems over flat fading channels with memory and incomplete side information," *Proc. IEEE Internat. Sympos. Inform. Theory*, p. 478, July 2002.

[32] T. Koch and A. Lapidoth, "The fading number and degrees of freedom in non-coherent MIMO fading channels: A peace pipe," *Proc. IEEE Internat. Sympos. Inform. Theory*, September 2005.

348

[33] G. B. Giannakis, Y. Hua, P. Stoica, and L. Tong, *Signal Processing Advances in Wireless and Mobile Communications: Trends in Single- and Multi-user Systems*, Prentice-Hall, New York, 2001.

[34] A. Molisch, M. Win, and J. H. Winters, "Reduced-complexity transmit/receive-diversity systems," *IEEE Trans. Signal Proc.*, pp. 2729–38, November 2003.

[35] H. Gamal, G. Caire, and M. Damon, "Lattice coding and decoding achieve the optimal diversity-multiplexing trade-off of MIMO channels," *IEEE Trans. Inform. Theory*, pp. 968–85, June 2004.

[36] V. Tarokh, N. Seshadri, and A. Calderbank, "Space-time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Trans. Inform. Theory*, pp. 744–65, March 1998.

[37] H. Yao and G. Wornell, "Structured space-time block codes with optimal diversity-multiplexing tradeoff and minimum delay," *Proc. IEEE Globecom Conf.*, pp. 1941–5, December 2003.

[38] L. Zheng and D. N. Tse, "Diversity and multiplexing: A fundamental trade-off in multiple antenna channels," *IEEE Trans. Inform. Theory*, pp. 1073–96, May 2003.

[39] M. O. Damen, H. El Gamal, and N. C. Beaulieu, "Linear threaded algebraic space-time constellations," *IEEE Trans. Inform. Theory*, pp. 2372–88, October 2003.

[40] H. El Gamal and M. O. Damen, "Universal space-time coding," *IEEE Trans. Inform. Theory*, pp. 1097–1119, May 2003.

[41] R. W. Heath, Jr., and A. J. Paulraj, "Switching between multiplexing and diversity based on constellation distance," *Proc. Allerton Conf. Commun., Control, Comput.*, pp. 212–21, October 2000.

[42] R. W. Heath, Jr., and D. J. Love, "Multi-mode antenna selection for spatial multiplexing with linear receivers," *IEEE Trans. Signal Proc.* (to appear).

[43] V. Jungnickel, T. Haustein, V. Pohl, and C. Von Helmolt, "Link adaptation in a multi-antenna system," *Proc. IEEE Veh. Tech. Conf.*, pp. 862–6, April 2003.

[44] J.-C. Guey, M. P. Fitz, M. Bell, and W.-Y. Kuo, "Signal design for transmitter diversity wireless communication systems over Rayleigh fading channels," *IEEE Trans. Commun.*, pp. 527–37, April 1999.

[45] V. Tarokh, A. Naguib, N. Seshadri, and A. Calderbank, "Space-time codes for high data rate wireless communication: Performance criteria in the presence of channel estimation errors, mobility, and multiple paths," *IEEE Trans. Commun.*, pp. 199–207, February 1999.

[46] S. Baro, G. Bauch, and A. Hansman, "Improved codes for space-time trellis coded modulation," *IEEE Commun. Lett.*, pp. 20–2, January 2000.

[47] J. Grimm, M. Fitz, and J. Korgmeier, "Further results in space-time coding for Rayleigh fading," *Proc. Allerton Conf. Commun., Control, Comput.*, pp. 391–400, September 1998.

[48] H. Gamal and A. Hammons, "On the design of algebraic space-time codes for MIMO block-fading channels," *IEEE Trans. Inform. Theory*, pp. 151–63, January 2003.

[49] A. Naguib, N. Seshadri, and A. Calderbank, "Increasing data rate over wireless channels," *IEEE Signal Proc. Mag.*, pp. 76–92, May 2000.

[50] V. Tarokh, H. Jafarkhani, and A. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Trans. Inform. Theory*, pp. 1456–67, July 1999.

[51] V. Gulati and K. R. Narayanan, "Concatenated codes for fading channels based on recursive space-time trellis codes," *IEEE Trans. Wireless Commun.*, pp. 118–28, January 2003.

[52] Y. Liu, M. P. Fitz, and O. Y. Takeshita, "Full-rate space-time codes," *IEEE J. Sel. Areas Commun.*, pp. 969–80, May 2001.

[53] K. R. Narayanan, "Turbo decoding of concatenated space-time codes," *Proc. Allerton Conf. Commun., Control, Comput.*, pp. 217–26, September 1999.

[54] B. Hochwald and T. Marzetta, "Unitary space-time modulation for multiple-antenna communications in Rayleigh flat fading," *IEEE Trans. Inform. Theory*, pp. 543–64, March 2000.

[55] D. Gesbert, M. Shafi, D.-S. Shiu, P. Smith, and A. Naguib, "From theory to practice: An overview of MIMO space-time coded wireless systems," *IEEE J. Sel. Areas Commun.*, pp. 281–302, April 2003. 349

[56] E. G. Larsson and P. Stoica, *Space-Time Block Coding for Wireless Communications*, Cambridge University Press, 2003.

[57] P. Wolniansky, G. Foschini, G. Golden, and R. Valenzuela, "V-blast: An architecture for realizing very high data rates over the rich-scattering wireless channel," *Proc. URSI Internat. Sympos. Signal Syst. Elec.*, pp. 295–300, October 1998.

[58] G. Foschini, G. Golden, R. Valenzuela, and P. Wolniansky, "Simplified processing for high spectral efficiency wireless communication employing multi-element arrays," *IEEE J. Sel. Areas Commun.*, pp. 1841–52, November 1999.

[59] G. Bauch and A. Naguib, "Map equalization of space-time coded signals over frequency selective channels," *Proc. IEEE Wireless Commun. Network Conf.*, pp. 261–5, September 1999.

[60] C. Fragouli, N. Al-Dhahir, and S. Diggavi, "Pre-filtered space-time M-BCJR equalizer for frequency selective channels," *IEEE Trans. Commun.*, pp. 742–53, May 2002.

[61] A. Naguib, "Equalization of transmit diversity space-time coded signals," *Proc. IEEE Globecom Conf.*, pp. 1077–82, December 2000.

[62] J. Winters, "Smart antennas for wireless systems," *IEEE Pers. Commun. Mag.*, pp. 23–7, February 1998. 350



第11章 均 衡

第6章已经讲到，时延扩展会引起码间干扰（ISI），当时延扩展和码元周期可比时，ISI会引起本底误码。可以通过信号处理的方法来对抗ISI，均衡就是指在接收端所采取的抗ISI技术。也可以在发送进行信号处理，使所发送的信号不容易受到时延扩展的影响，例如扩频和多载波调制。本章主要讨论均衡，第12、第13章分别介绍多载波调制和扩频技术。

当码元周期T_s与信道的均方根时延扩展σ_{T_m}可比时，就需要考虑ISI问题。例如，无绳电话通常在室内使用，时延扩展很小，语音的数据率又很低，所以不必考虑均衡。但IS-136数字蜂窝电话通常用在室外，其$\sigma_{T_m} \approx T_s$，此时均衡就是必须的。高速数据业务对时延扩展更为敏感，需要高性能的均衡器或其他可以减小ISI的技术。事实上如何对抗时延扩展带来的影响是高速无线数据传输所面临的重要挑战之一。

经过均衡器的同时有信号和噪声，噪声经过均衡器后有可能被放大。因此，均衡器的设计需要在消除ISI和噪声增强之间取得平衡。非线性均衡中噪声增强的问题相对小一些，但其复杂度较高，下面会详细讨论。另外，均衡器需要估计信道的冲激响应或频率响应。无线信道是时变的，所以均衡器首先估计出信道的频率或冲激响应（训练），之后随着信道的变化随时更新估计值（跟踪）。能够进行训练和跟踪的均衡器可以自动适应信道的变化，称为自适应均衡。当信道变化过快时，均衡器的训练和跟踪会变得非常困难。本章讨论均衡器设计方面的一些问题，包括ISI抵消和噪声增强之间的平衡问题、线性和非线性的均衡器的设计问题及其性能、均衡器的训练和跟踪过程等。

351

均衡器可以在基带、射频或中频实现。由于数字滤波器体积小、价格低、容易调整、功耗小，所以多数情况下，均衡器是经过模/数转换后以数字化的方式实现的。本章主要考虑数字均衡器。但为了简单起见，下节用模拟均衡器来解释噪声增强。

11.1 均衡器中的噪声增强

均衡器的目的是消除ISI的影响。但在消除ISI的同时，必须要顾及到噪声功率的增强问题。我们用图11-1这个简单的模拟均衡器来说明设计中不考虑噪声影响，只是单纯消除ISI时存在的问题。假设信号$s(t)$经过了一个频率响应为$H(f)$的信道，在接收机的前端叠加了高斯白噪声$n(t)$，从而输入信号为$Y(f)=S(f)H(f)+N(f)$，其中$N(f)$是功率谱密度为$N_0/2$的白噪声。若信号带宽为B，那么信号带宽内的噪声功率为N_0B。如欲完全消除信道中引入的ISI，只需在接收端引入这样一个模拟均衡器：

$$H_{eq}(f)=1/H(f) \tag{11-1}$$

经过均衡之后的接收信号$Y(f)$为

$$[S(f)H(f)+N(f)]H_{eq}(f)=S(f)+N'(f)$$

其中$N'(f)$是有色高斯噪声，其功率谱密度为$0.5N_0/|H(f)|^2$。上式表明所有ISI都被消除了。

但是，若在$s(t)$的带宽范围内$H(f)$有零点，即若对于某一f_0，有$H(f_0)=0$，那么噪声$N'(f)$的功率将为无限大。即使频谱没有零点，如果某些频率处有很大的衰减，那么均衡器$H_{eq}(f)=1/H(f)$也会使这些频率上的噪声显著增大。在这种情况下，虽然ISI被消除了，但因为信噪比大大降低，所以性能也会很差。因此，均衡设计应当在减小ISI的同时最大化均衡器输出的信噪比。一般来说，数字的线性均衡器其原理也近似是将信道的频率响应进行反转，噪声增强的问题较大，而非线性均衡器并不是将信道进行反转，因此噪声增强的问题较小。下一节将概述不同类型的线性和非线性均衡器，包括均衡器结构以及训练跟踪过程中的抽头系数更新算法。

352

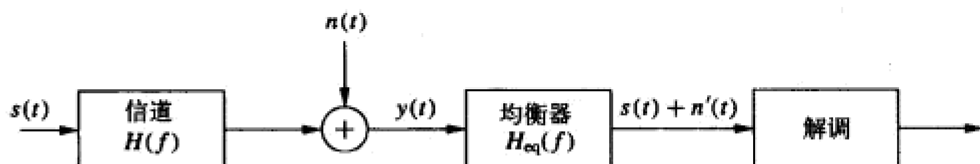


图11-1 通过模拟均衡器说明噪声增强

例11.1：设信道的频率响应为$H(f)=1/\sqrt{|f|}, |f|<B$，B为信道带宽。给定噪声的功率谱密度为$N_0/2$，B=30kHz。请问，采用和不采用信道反转式均衡器时，噪声功率分别为多少？

解：不采用均衡器时的噪声功率为$N_0B=3N_0\times10^4$。采用均衡器后，噪声功率谱密度变成$0.5N_0|H_{eq}(f)|^2=0.5N_0/|H(f)|^2=0.5|f|N_0, |f|<B$。因此噪声功率为

$$0.5N_0\int_{-B}^{B}|f|\,\mathrm{d}f=0.5N_0B^2=4.5N_0\times10^8$$

噪声功率增加了4个数量级！

11.2 均衡器的类型

均衡技术主要分为线性和非线性两种类型。线性均衡实现简单、易于理解，但多数无线通信系统并没有采用线性均衡，因为它的噪声增强要比非线性均衡大。最常用的非线性均衡是实现简单、性能也不错的判决反馈均衡（decision-feedback equalization，DFE）。但在低信噪比时，DFE存在误码传播的问题，进而会导致性能恶化。最优的均衡技术是最大似然序列估计（maximum likelihood sequence estimation，MLSE），但其复杂度随时延扩展成指数增长，这对多数信道来说是难以实用的。人们经常将MLSE作为比较各种均衡技术的性能上界。图11-2概括了不同的均衡器类型、结构和抽头更新算法。

均衡器又可分为逐符号（symbol-by-symbol，SBS）均衡或序列估计（sequence estimator，SE）均衡。逐符号均衡器对每个符号单独进行ISI消除并进行检测，图11-2中所有的线性均衡器和反馈判决均衡器都是逐符号均衡。序列估计则是检测整个符号序列，ISI的影响也是检测过程的一部分。MLSE是序列检测的最优形式，但复杂度很高。

线性或非线性均衡器一般采用横向滤波器或格形滤波器。横向滤波器有$N-1$个延时单元、N个可调的复数抽头。格形滤波器采用了更复杂的递归结构[1]，它有更好的数

值稳定性和收敛性，并且也便于改变结构长度[2]。本章主要考虑横向结构，格形结构及其性能见文献[1~4]。

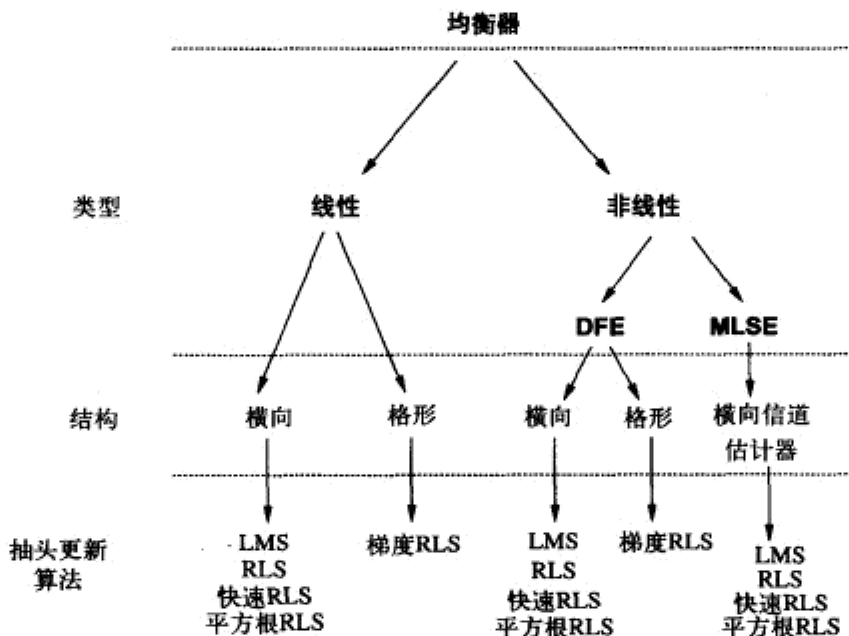


图11-2 均衡器的类型、结构和算法

除了类型和结构之外，自适应均衡器还需要一个在训练和跟踪过程中更新滤波器的抽头系数的算法。多年来，人们已经提出了很多这样的算法。这些算法一般需要综合考虑复杂性、收敛速度以及稳定性。

353

本章随后的内容将先讨论无ISI传输条件，然后具体讨论不同的均衡器类型、它们的结构以及更新算法。本章关于均衡器的分析以带通系统的等效基带表示为基础，详见附录A。

11.3 折叠谱和无ISI传输

均衡器通常以数字方式实现，图11-3是采用数字均衡器时，系统端到端的等效基带框图。输入符号d_k先经过脉冲成形滤波器$g(t)$，然后在等效基带冲激响应为$c(t)$的ISI信道上传输。若定义组合信道的冲激响应为$h(t)\triangleq g(t)*c(t)$，那么信息序列$d(t)=\sum_k d_k\delta(t-kT_s)$经过信道后为$d(t)*g(t)*c(t)$。我们在5.5节中已经讲过，脉冲成形$g(t)$可用来改善发送信号的频谱性能，它是可以设计的，但信道$c(t)$是自然引入的，不受设计者的控制。

354

在接收机的前端，接收信号叠加了单边功率谱密度为N_0的等效低通白高斯噪声$n(t)$，成为$w(t)$。此信号经过模拟匹配滤波器$g_m^*(-t)$得到等效基带输出$y(t)$，然后由模/数转换器抽样为数字信号。匹配滤波器是为了使抽样时刻以及后续处理中的信噪比最大化[1]。由5.1节可知，在加性白高斯噪声干扰下，采用与脉冲形状匹配的滤波器可以最大化抽样信号的信噪比。因此，对于图11-3所示的系统，为了使抽样信噪比最大，应当将$w(t)$经过与$h(t)$相匹配的滤波器，即理想情况应当是$g_m(t)=h(t)$。但是，信道冲激响应$c(t)$是时

变的，模拟滤波器又不容易调整，所以一般很难做到$g_m(t)=h(t)$。因此，如何选择$g_m(t)$使系统具有良好性能也是均衡器设计中的一个重要考虑。通常设计$g_m(t)$与脉冲$g(t)$匹配，这样的设计在$c(t)=\delta(t)$时是最优的，但在$c(t)\neq\delta(t)$时显然是次优的。$g_m(t)$与$h(t)$不匹配会使系统性能大大降低，也使接收端对定时误差更为敏感。若以比符号速率大得多的速率对$y(t)$进行抽样，再对这个抽样信号设计相应的均衡器，可使问题有所解决。这种情形称为分数抽头均衡（fractionally spaced equalization），详见[4]的10.2.4节。

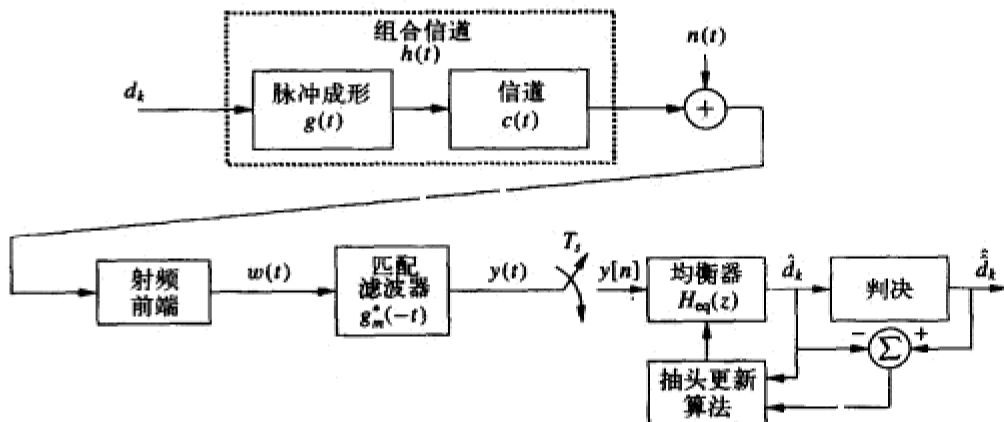


图11-3 端到端的等效基带系统

均衡器的输出是对发送符号的估计，这个估计值被判决为一个可能的发送符号。在训练阶段，抽头更新算法用均衡器的输出来更新抽头值，以使均衡器的输出和已知的训练序列相匹配。在跟踪期间，则是用符号的判决误差来调整均衡器的系数。

令$f(t)$表示发送端脉冲成形、信道和匹配滤波器冲激响应合成的等效低通冲激响应：

$$f(t)\triangleq g(t)*c(t)*g_m^*(-t) \tag{11-2}$$

那么匹配滤波器输出为

$$y(t)=d(t)*f(t)+n_g(t)=\sum d_k f(t-kT_s)+n_g(t) \tag{11-3}$$

355

其中$n_g(t)=n(t)*g_m^*(-t)$是均衡器输入端的等效低通噪声，T_s是码元周期。令$f[n]=f(nT_s)$表示$f(t)$的抽样、$y[n]=y(nT_s)$表示$y(t)$的抽样，则

$$\begin{aligned}y[n]&=\sum_{k=-\infty}^{\infty}d_k f(nT_s-kT_s)+n_g(nT_s)\\&=\sum_{k=-\infty}^{\infty}d_k f[n-k]+v[n]\\&=d_n f[0]+\sum_{k\neq n}d_k f[n-k]+v[n]\end{aligned} \tag{11-4}$$

其中第一项为期望的数据比特，第二项为ISI，第三项为噪声的抽样。由式（11-4）

1. 虽然匹配滤波器也可以抽样后用数字的方法实现，但抽样前的模拟匹配滤波可以降低抽样器的动态范围，从而能大大降低成本。

可知无ISI的条件是：对$k\neq n$有$f[n-k]=0$，也即$f[k]=\delta[k]f[0]$。此时式（11-4）成为$y[n]=d_n f[0]+v[n]$。

我们现在来证明，满足无ISI传输，即$f[k]=\delta[k]f[0]$的充分必要条件是：

$$F_\Sigma(f)\triangleq\frac{1}{T_s}\sum_{n=-\infty}^{\infty}F\left(f+\frac{n}{T_s}\right)=f[0] \tag{11-5}$$

这里的$F_\Sigma(f)$是周期为$1/T_s$的周期函数，常称为折叠谱，$F_\Sigma(f)=f[0]$表明折叠谱是平坦的。

为了证明无ISI传输和平坦折叠谱之间的等价性，首先注意到

$$\begin{aligned}f(k)=f(kT_s)&=\int_{-\infty}^{\infty}F(f)\mathrm{e}^{\mathrm{j}2\pi fkT_s}\mathrm{d}f\\&=\sum_{n=-\infty}^{\infty}\int_{0.5(2n-1)/T_s}^{0.5(2n+1)/T_s}F(f)\mathrm{e}^{\mathrm{j}2\pi fkT_s}\mathrm{d}f\\&=\sum_{n=-\infty}^{\infty}\int_{-0.5/T_s}^{0.5/T_s}F\left(f'+\frac{n}{T_s}\right)\mathrm{e}^{\mathrm{j}2\pi(f'+n/T_s)kT_s}\mathrm{d}f'\\&=\int_{-0.5/T_s}^{0.5/T_s}\mathrm{e}^{\mathrm{j}2\pi fkT_s}\left[\sum_{n=-\infty}^{\infty}F\left(f+\frac{n}{T_s}\right)\right]\mathrm{d}f\end{aligned} \tag{11-6}$$

式（11-6）表明$F_\Sigma(f)$的傅里叶级数为

$$F_\Sigma(f)=\frac{1}{T_s}\sum_k f[k]\mathrm{e}^{-\mathrm{j}2\pi fkT_s}$$

先证明平坦折叠谱表明$f[k]=\delta[k]f[0]$。假设式（11-5）成立，将其代入式（11-6）可得

$$f(k)=T_s\int_{-0.5/T_s}^{0.5/T_s}\mathrm{e}^{\mathrm{j}2\pi fkT_s}f[0]\mathrm{d}f=\frac{\sin\pi k}{\pi k}f[0]=\delta[k]f[0] \tag{11-7}$$

即为所求。再来证明$f[k]=\delta[k]f[0]$表明折叠谱平坦。由（11-6）及$F_\Sigma(f)$的定义可得

356

$$f[k]=T_s\int_{-0.5/T_s}^{0.5/T_s}F_\Sigma(f)\,\mathrm{e}^{\mathrm{j}2\pi fkT_s}\mathrm{d}f \tag{11-8}$$

即$f[k]$是$F_\Sigma(f)$的傅里叶反变换'。因此，若$f[k]=\delta[k]f[0]$，必然$F_\Sigma[f]=f[0]$。

例11.2：假设组合信道的等效低通冲激响应为$f(t)=\mathrm{sinc}(t/T_s)$，求折叠谱，并判断信道是否有ISI。

解：$f(t)$的傅里叶变换为

$$F(f)=T_s\mathrm{rect}(fT_s)=\begin{cases}T_s & |f|<0.5/T_s\\0.5T_s & |f|=0.5/T_s\\0 & |f|>0.5/T_s\end{cases}$$

因此

$$F_\Sigma(f)=\frac{1}{T_s}\sum_{n=-\infty}^{\infty}F\left(f+\frac{n}{T_s}\right)=1$$

PDG

由于折叠谱是平坦的，所以没有ISI。也可以由

$$f(nT_s) = \text{sinc}\left(\frac{nT_s}{T_s}\right) = \text{sinc}(n) = \begin{cases} 1 & n = 0 \\ 0 & n \neq 0 \end{cases}$$

得到$f[k]=\delta[k]$，它是和折叠谱平坦等价的无ISI条件。

11.4 线性均衡

如果$F_\Sigma(f)$不平，可以用图11-3中的均衡器$H_{eq}(z)$来减小ISI。本节假设均衡器为一个$N=2L+1$抽头的横向滤波器

$$H_{eq}(z) = \sum_{i=-L}^{L} w_i z^{-i} \tag{11-9}$$

均衡器的长度N一般是根据一些实现上的考虑来确定的，N越大，则复杂度越高、延时越大。对于因果线性均衡器，有$w_i=0$，$i<0$。给定N后，均衡器的设计就是要给出：(i) 给定信道频率响应时，各抽头系数的值$\{w_i\}_{i=-L}^{L}$；(ii) 信道变化时，这些抽头系数的更新算法。无线通信系统的基本性能指标是差错率（或中断率），对于给定的信道，最佳的均衡器系数应该能使差错率最小。但是按这个准则来最优化$\{w_i\}$是极其困难的，因此我们只能通过间接的优化来减小ISI，并尽量避免出现前面模拟均衡中所提到的那种噪声增强。下面将讨论两种线性均衡器：迫零（zero-forcing，ZF）均衡器和最小均方误差（minimum mean-square error，MMSE）均衡器。前者可以完全消除ISI，但是噪声增强比较显著。后者兼顾了降低ISI和避免噪声增强的问题，它能最小化发送符号和均衡器输出符号之间的均方误差。因此，MMSE均衡的错误率性能要好于ZF均衡。 357

11.4.1 迫零均衡

由（11-4）可知，均衡器的输入$\{y_n\}$可以用组合信道的等效低通冲激响应$f(t)=h(t)*g^*(-t)$的抽样来表示，即

$$Y(z) = D(z)F(z) + N_g(z) \tag{11-10}$$

其中$N_g(z)$是匹配滤波器$G^*(1/z^*)$输出的噪声样值的z变换。$F(z)$为

$$F(z) = H(z)G_m^*\left(\frac{1}{z^*}\right) = \sum_n f(nT_s)z^{-n} \tag{11-11}$$

迫零均衡器将消除由$f(t)$引入的所有ISI。由式（11-10）可知，能实现这一目的的均衡器为

$$H_{ZF}(z) = \frac{1}{F(z)} \tag{11-12}$$

它是前述的模拟均衡器式（11-1）在时间上离散化后的等效基带均衡器，同样也有噪声

1. 也可用$F_\Sigma(f)$的傅里叶级数表达式证明这一点。

PDG

增强的问题。具体而言，均衡器输出的噪声样值的功率谱密度为

$$\begin{aligned} N(z) &= N_0|G_m^*(1/z^*)|^2|H_{ZF}(z)|^2 = \frac{N_0|G_m^*(1/z^*)|^2}{|F(z)|^2} \\ &= \frac{N_0|G_m^*(1/z^*)|^2}{|H(z)|^2|G_m^*(1/z^*)|^2} = \frac{N_0}{|H(z)|^2} \end{aligned} \tag{11-13}$$

可见，如果信道$H(z)$在信号带宽范围内的任一频率处受到严重衰减（这在频率选择衰落信道中是很常见的），噪声功率就会显著增加。正是这一点使人们认识到，均衡器的设计必须要在消除ISI和噪声增强之间寻求一种最佳的平衡。能达到这种平衡的均衡器之一就是下一节要讨论的MMSE均衡。

358

由$H_{ZF}(z)=1/F(z)$所定义的迫零均衡器不一定能用有限冲激响应（finite-impulse response，FIR）滤波器来实现。也就是说，未必能找出有限个系数$w_{-L},\cdots w_L$使得

$$w_{-L}z^L+\cdots+w_Lz^{-L}=\frac{1}{F(z)} \tag{11-14}$$

在这种情况下，我们需要找出一组系数$\{w_i\}$，它能最佳地近似迫零均衡。注意这个近似必须对所有z值都有效，所以求解并非易事。有很多方法可以求解这种近似，其中之一是将$H_{ZF}(z)$表示为无限冲激响应（IIR）：$1/F(z)=\sum_{i=-\infty}^{\infty}c_iz^{-i}$，然后取$w_i=c_i$。可以证明，当$z=e^{j\omega}$时，这种方法可以使下式最小：

$$\left|\frac{1}{F(z)}-(w_{-L}z^L+\cdots+w_Lz^{-L})\right|^2$$

除此之外，还可以寻求能使峰值畸变（最坏情况下的ISI）最小化的抽头系数。寻找最小峰值畸变的抽头系数是一个凸优化问题，可用一些标准解法来求解，如最速下降法[4]。

例11.3：假设信道冲激响应为

$$h(t)=\begin{cases} e^{-t/\tau} & t\geqslant 0 \\ 0 & \text{其他} \end{cases}$$

求此信道的二抽头迫零均衡器。

解：我们有

$$h[n]=1+e^{-T_s/\tau}\delta[n-1]+e^{-2T_s/\tau}\delta[n-2]+\cdots.$$

因此

$$\begin{aligned} H(z) &= 1+e^{-T_s/\tau}z^{-1}+e^{-2T_s/\tau}z^{-2}+e^{-3T_s/\tau}z^{-3}+\cdots \\ &= \sum_{n=0}^{\infty}(e^{-T_s/\tau}z^{-1})^n=\frac{z}{z-e^{-T_s/\tau}} \end{aligned}$$

因此$H_{eq}(z)=1/H(z)=1-e^{-T_s/\tau}z^{-1}$。因此所求两抽头均衡器的抽头系数为$w_0=1, w_1=e^{-T_s/\tau}$。

11.4.2 最小均方误差均衡

MMSE均衡的目标是最小化发送符号d_k和均衡器输出 $\hat{d}_k$ 之间的均方误差（mean-square error，MSE），也就是要选择$\{w_i\}$使 $E\left[d_k-\hat{d}_k\right]^2$ 最小。MMSE均衡是线性均衡器，其输出 $\hat{d}_k$ 是输入样值$y[k]$的线性组合：

$$\hat{d}_k=\sum_{i=-L}^{L} w_i y[k-i] \tag{11-15}$$

这样，求最优系数$\{w_i\}$的问题是一个标准的线性估计问题。若均衡器输入的噪声是白噪声，这就是一个标准的维纳滤波的问题。但由于接收前端存在的匹配滤波器 $g_m^*(-t)$，输入到均衡器的噪声是有色噪声，功率谱为 $N_0\left|G_m^*\left(1/z^*\right)\right|^2$。因此，为了利用已有的技术，我们将滤波器$H_{eq}(z)$分为两部分：噪声白化部分 $1/G_m^*\left(1/z^*\right)$和消除ISI部分 $\hat{H}_{eq}(z)$，如图11-4所示。

359

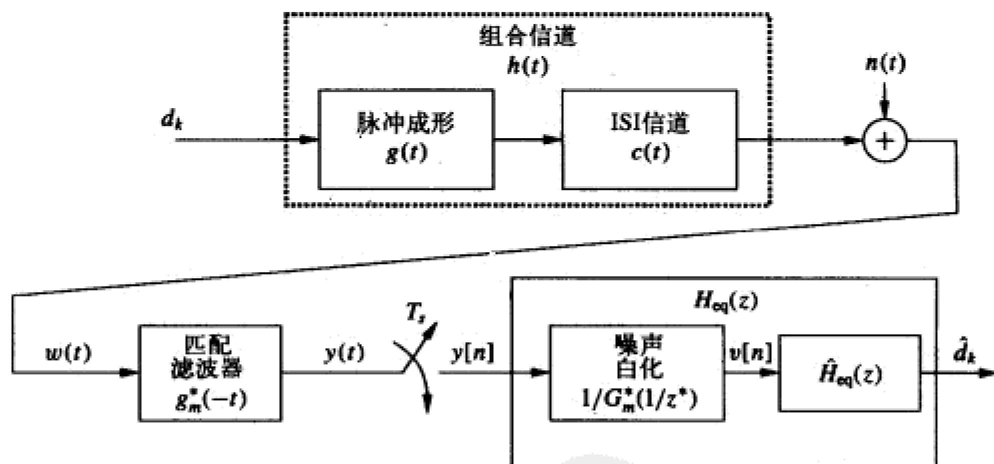


图11-4 带有噪声白化滤波的MMSE均衡

顾名思义，噪声白化滤波器的目的就为了使输出噪声的功率谱成为常数。由于输入的噪声功率谱为 $N_0\left|G_m^*\left(1/z^*\right)\right|^2$，因此近似的噪声白化滤波器为 $1/G_m^*\left(1/z^*\right)$。这样，输出的噪声功率谱变成 $N_0\left|G_m^*\left(1/z^*\right)\right|^2\Big/\left|G_m^*\left(1/z^*\right)\right|^2=N_0$。注意滤波器 $1/G_m^*\left(1/z^*\right)$并非唯一能够实现白化的滤波器，根据其他考虑（如稳定性），还可以选择其他的噪声白化滤波器。有一点看起来有些奇怪，就是我们先引入了一个匹配滤波器 $g_m^*(-t)$，而后又用白化滤波抵消了匹配滤波。需要注意的是，我们用匹配滤波器是为了最大化抽样的信噪比，再通过噪声白化去除匹配滤波器的影响之后，我们就可以简化MMSE滤波器 $\hat{H}_{eq}(z)$ 的设计。实际上，如果噪声白化滤波器性能没有达到最优的话，其影响可以被 $\hat{H}_{eq}(z)$ 滤波器的设

计消除，如下面要讲到的IIR MMSE均衡。

假设滤波器 $\hat{H}_{eq}(z)$ 的输入为v_n，有$N=2L+1$个抽头

$$\hat{H}_{eq}(z)=\sum_{i=-L}^{L} w_i z^{-i} \tag{11-16}$$

我们的目标是设计滤波器系数$\{w_i\}$使 $E\left[d_k-\hat{d}_k\right]^2$ 最小化，这同样也是滤波器$H_{eq}(z)$的目标，我们只是增加了一个噪声白化滤波器以使系数的求解问题得到简化。记滤波器 $\hat{H}_{eq}(z)$ 的输入向量为 $v^T=(v[k+L],v[k+L-1],\cdots,v[k-L])=(v_{k+L},v_{k+L-1},\cdots,v_{k-L})$，这个向量产生了滤波器输出 $\hat{d}_k$。令向量$w^T=(w_{-L},\cdots,w_L)$表示滤波器系数，则有

$$\hat{d}_k=w^T v=v^T w \tag{11-17}$$

我们需要最小化均方误差

$$J=E[d_k-\hat{d}_k]^2=E[w^T vv^H w^*-2\mathrm{Re}\{v^H w^* d_k\}+|d_k|^2] \tag{11-18}$$

令$M_v=E[vv^H]$，$v_d=E[v^H d_k]$，则M_v是一个$N\times N$的复共扼矩阵，v_d是长为N的行向量。假设$E[d_k]^2=1$，则均方误差J为

360

$$J=w^T M_v w^*-2\mathrm{Re}\{v_d w^*\}+1 \tag{11-19}$$

令梯度 $\nabla_w J=0$ 可求解出最优抽头向量w。由式（11-19）可得梯度为

$$\nabla_w J=\left(\frac{\partial J}{\partial w_{-L}},\cdots,\frac{\partial J}{\partial w_L}\right)=2w^T M_v-2v_d \tag{11-20}$$

令其为零得$w^T M_v=v_d$，因为最优抽头系数为

$$w_{opt}=(M_v^T)^{-1}v_d^T \tag{11-21}$$

注意求出w_{opt}需要矩阵求逆，其运算量是非常大的，通常为N^2到N^3的数量级。代入这些最优抽头系数可得最小均方误差为

$$J_{min}=1-v_d M_v^{-1} v_d^H \tag{11-22}$$

对于无限长均衡器，$v^T=(v_{n+\infty},\cdots v_n,\cdots v_{n-\infty})$，$w^T=(w_{-\infty},\cdots,w_0\cdots,w_\infty)$，于是$w^T M_v=v_d$可写为（见文献[5]的7.4节）

$$\sum_{i=-\infty}^{\infty} w_i(f[j-i]+N_0\delta[j-i])=g_m^*[-j] \quad -\infty\leqslant j\leqslant\infty \tag{11-23}$$

等式两端进行z变换，并注意到 $\hat{H}_{eq}(z)$ 是滤波器系数w的z变换，可得

$$\hat{H}_{eq}(z)(F(z)+N_0)=G_m^*(1/z^*) \tag{11-24}$$

由此解出 $\hat{H}_{eq}(z)$ 为

$$\hat{H}_{eq}(z) = \frac{G_m^*(1/z^*)}{F(z) + N_0} \tag{11-25}$$

MMSE均衡器包括噪声白化滤波器 $1/G_m^*(1/z^*)$ 和消除ISI的滤波器 $\hat{H}_{eq}(z)$，所以滤波器长度无限时的完整MMSE均衡器可以表示为

$$H_{eq}(z) = \frac{\hat{H}_{eq}(z)}{G_m^*(1/z^*)} = \frac{1}{F(z) + N_0} \tag{11-26}$$

361

这个结果有三个有意思的地方值得注意。首先，这个理想的无限长MMSE均衡器抵消了噪声白化滤波器。其次，除了噪声项N_0外，此滤波器与迫零滤波器一致，因此无噪声时二者是等价的。最后，这个理想的均衡器设计显然达到了信道反转和噪声增强之间的平衡：如果$F(z)$在某一频率严重衰减，分母中的噪声项N_0的存在防止了输出噪声的显著提高。而在噪声功率密度N_0相对于$F(z)$较小的频率处，均衡器实际上是反转了$F(z)$。

对于（11-26）给出的均衡器，文献[5]的7.4.1节指出最小的均方误差（11-22）可以用折叠谱$F_\Sigma(f)$来表示：

$$J_{\min} = T_s\int_{-0.5/T_s}^{0.5/T_s} \frac{N_0}{F_\Sigma(f) + N_0} df \tag{11-27}$$

上式有这样一些特性。首先，可以证明$0 \leqslant J_{\min} = E\left[d_k - \hat{d}_k\right]^2 \leqslant 1$。此外，只要在信号带宽范围内$F_\Sigma(f) \neq 0$，且没有噪声（$N_0=0$），那么$J_{\min}=0$。另外，当$N_0=\infty$时$J_{\min}=1$。

例11.4：若折叠谱$F_\Sigma(f)$平坦且$F_\Sigma(f)=f[0]$，求信噪比趋于无限大和无限小时的$J_{\min}$。

解：如果 $F_\Sigma(f) = f[0] \triangleq f_0$，那么

$$J_{\min} = T_s\int_{-0.5/T_s}^{0.5/T_s} \frac{N_0}{f_0 + N_0} df = \frac{N_0}{f_0 + N_0}$$

当信噪比趋于无限大时，$f_0 >> N_0$，因此$J_{\min} \approx N_0/f_0 = N_0/E_s$，其中$E_s/N_0$为符号信噪比。当信噪比趋于无限小时，$N_0 >> f_0$，因此$J_{\min} = N_0/(N_0+f_0) \approx N_0/N_0 = 1$。

11.5 最大似然序列估计

最大似然序列估计对发送的符号序列进行估计，由于它不使用滤波器，因此没有噪声增强的问题。MLSE的结构同图11-3，只是将均衡器$H_{eq}(z)$和判决器换成了MLSE算法。当脉冲成形和信道响应组合的$h(t)$给定时，MLSE算法根据接收信号$w(t)$，选择出具有最大似然值的输入序列$\{d_k\}$。下面我们详细证明一下这个定理。

利用格莱姆-施密特正则化，在时间间隔$[0, LT_s]$内，$w(t)$可以表示成

$$w(t) = \sum_{n=1}^{N} w_n \phi_n(t) \tag{11-28}$$

其中$\{\phi_n(t)\}$是一组完备的正则集函数。在$[0, LT_s]$内，由于$w(t)$取决于$d_0, \cdots, d_L$，所以N与信道的记忆有关。由此我们可以得到

$$w_n = \sum_{k=-\infty}^{\infty} d_k h_{nk} + \nu_n = \sum_{k=0}^{L} d_k h_{nk} + \nu_n \tag{11-29}$$

其中

$$h_{nk} = \int_0^{LT_s} h(t - kT_s)\phi_n^*(t)\,\mathrm{d}t \tag{11-30}$$

362

$$\nu_n = \int_0^{LT_s} n(t)\phi_n^*(t)\,\mathrm{d}t \tag{11-31}$$

复高斯随机变量ν_n的零均值为零，方差为$0.5E[\nu_n^*\nu_m] = N_0\delta[n-m]$。因此，$w^N=(w_1, \cdots, w_N)$为多维高斯分布

$$p(w^N \mid d^L, h(t)) = \prod_{n=1}^{N}\left[\frac{1}{\pi N_0}\exp\left[-\frac{1}{N_0}\left|w_n - \sum_{k=0}^{L} d_k h_{nk}\right|^2\right]\right] \tag{11-32}$$

给定接收信号$w(t)$或其等价向量w^N，MLSE将其译为符号序列d^L，此序列能使似然函数$p(w^N|d^L, h(t))$或其对数最大化，即MLSE的输出序列为

$$\begin{aligned}
\hat{d}^L &= \arg\max[\log p(w^N \mid d^L, h(t))] \\
&= \arg\max\left[-\sum_{n=1}^{N}\left|w_n - \sum_k d_k h_{nk}\right|^2\right] \\
&= \arg\max\left[-\sum_{n=1}^{N}|w_n|^2 + \sum_{n=1}^{N}\left(w_n^*\sum_k d_k h_{nk} + w_n\sum_k d_k^* h_{nk}^*\right)\right. \\
&\qquad \left. - \sum_{n=1}^{N}\left(\sum_k d_k h_{nk}\right)\left(\sum_m d_m^* h_{nm}^*\right)\right] \\
&= \arg\max\left[2\,\mathrm{Re}\left\{\sum_k d_k^* \sum_{n=1}^{N} w_n h_{nk}^*\right\} - \sum_k\sum_m d_k d_m^* \sum_{n=1}^{N} h_{nk}h_{nm}^*\right]
\end{aligned} \tag{11-33}$$

注意到

$$\sum_{n=1}^{N} w_n h_{nk}^* = \int_{-\infty}^{\infty} w(\tau)h^*(\tau - kT_s)\,\mathrm{d}\tau = y[k] \tag{11-34}$$

$$\sum_{n=1}^{N} h_{nk}h_{nm}^* = \int_{-\infty}^{\infty} h(\tau - kT_s)h^*(\tau - mT_s)\,\mathrm{d}\tau = f[k-m] \tag{11-35}$$

结合式（11-33）、式（11-34）和式（11-35）可得

$$\hat{d}^L = \arg\max\left[2\,\mathrm{Re}\left\{\sum_k d_k^* y[k]\right\} - \sum_k\sum_m d_k d_m^* u[k-m]\right] \qquad (11\text{-}36)$$

由此可见，MLSE的输出只与抽样值$\{y[k]\}$和信道参数$u[n-k]=u\ (nT_s-kT_s)$有关，其中$u(t)=h\ (t)*h^*(-t)$。由于MLSE是从匹配滤波之前的信道输出$w(t)$为基础导出的，说明对MSLE检测来说，图11-3中匹配滤波器对$g_m(t)=h\ (t)$时是最佳的。（通常匹配滤波器对于加性高斯白噪声下的信号检测是最优的。上述讨论表明，采用MLSE检测时，即使有ISI，匹配滤波器也是最优的）。

363

可以用维特比算法来降低MLSE的复杂度[4-7]。不过维特比均衡的复杂度仍然随信道时延扩展成指数增长。一种可以大大降低复杂度的非线性均衡技术是判决反馈均衡。

11.6 判决反馈均衡

判决反馈均衡包括一个类似于线性均衡的前馈滤波器$W(z)$和一个反馈滤波器$V(z)$，如图11-5所示。前馈滤波器的输入是接收序列，反馈滤波器的输入是已判决的序列。判决反馈均衡器将已检测的符号$\left\{\hat{\hat{d}}_k\right\}$通过反馈滤波器，来近似前馈滤波器$W(z)$和组合信道$F(z)$卷积后的ISI，从而能在接收符号中减掉ISI。由于$V(z)$处于反馈环中，因此它必须是严格因果的，否则系统将变得不稳定。反馈滤波器的作用是近似信道的频率响应，而不是对其进行反转，所以不存在噪声增强的问题。当信道的频谱有很多深衰落零点时，判决反馈均衡的性能一般比线性均衡好很多。

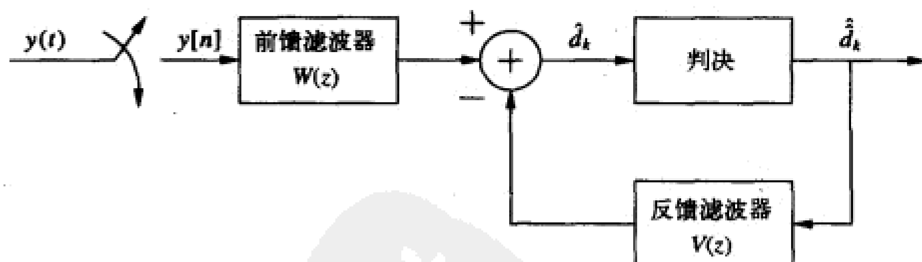


图11-5 判决反馈均衡器的结构

假设$W(z)$有N_1+1个抽头，$V(z)$有N_2个抽头，那么判决反馈均衡的输出可表示为

$$\hat{d}_k = \sum_{i=-N_1}^{0} w_i y[k-i] - \sum_{i=1}^{N_2} v_i \hat{\hat{d}}_{k-i}$$

选择$W(z)$和$V(z)$系数的常用准则是迫零（消除所有ISI）或者MMSE（最小化判决反馈均衡输出和原始符号之间的均方误差）。假设接收端的匹配滤波器与信道理想匹配于：$g_m(t)=h\ (t)$。为了获得因果滤波器$V(z)$，我们需要频谱的因式分解$H(z)G_m^*(1/z^*)=H(z)H^*(1/z^*)$（见文献[4]的10.1.2节）。也即要分解为 $H(z)H^*(1/z^*)=B(z)B^*(1/z^*)=\lambda^2 B_1(z)B_1^*(1/z^*)$，

其中$B(z)$是一个因果滤波器，$B_1(z)$则是将首项的系数归一化为1。当$W(z)$和$V(z)$为无限长时，文献[8]指出，用这种分解的方法按迫零准则得到的最优的前馈和反馈滤波器分别为$W(z)=1/\lambda^2 B_1^*(1/z^*)$ 和$V(z)=1-B_1(z)$。

364

对于MMSE准则，我们希望最小化 $E[d_k-\hat{d}_k]^2$ 。令$b_n=b[n]$表示上述因式分解$H(z)H^*(1/z^*)=B(z)B^*(1/z^*)$得到的$B(z)$的逆z变换，那么前馈滤波器的系数必满足如下的线性方程：

$$\sum_{i=-N_1}^{0} q_{li}\hat{w}_i = b_{-l}^*$$

其中 $q_{li}=\sum_{j=-l}^{0} b_j^* b_{j+l-i}+N_0\delta[l-i],\ i=-N_1,\cdots,0$，（见文献[4]的10.3.1节）。然后可用前馈滤波器系数决定反馈滤波器的系数为

$$v_k=-\sum_{i=-N_1}^{0} \hat{w}_i b_{k-i}$$

当没有判决错误，即 $\hat{\hat{d}}_k=d_k$ 时，这些系数可以完全消除ISI。Saltz证明[9]最终的最小均方误差为

$$J_{\min}=\exp\left[T_s\int_{-0.5/T_s}^{0.5/T_s}\ln\left[\frac{N_0}{F_\Sigma(f)+N_0}\right]\mathrm{d}f\right]$$

若反馈错误可以忽略的话，判决反馈均衡的最小均方误差通常要比线性均衡器的低得多。

若 $\hat{\hat{d}}_k\neq d_k$ ，则判决反馈均衡中存在反馈误差，此时通过反馈路径减掉的ISI并非与d_k对应的真正的ISI。这样的错误会传播到后面的比特判决中。因为ISI必须立即消除，不能允许任何译码延时，所以反馈路径所反馈的是译码前的编码比特，这就是说我们无法借助信道编码来避免误码传播现象。因此，在低信噪比时，误码传播将严重降低性能。如果允许在反馈路径上引入延时的话，可借助信道译码来解决这一问题[10]，也可采用下一节介绍的Turbo均衡。关于有编码的判决反馈均衡的系统描述可见文献[11~12]。判决反馈均衡的结构也可以推广到MIMO信道中[13]。

11.7　其他均衡方法

虽然MLSE是最佳均衡，但其复杂度使它未能广泛应用。人们在降低MLSE的复杂度方面已经作了许多工作（见文献[4]的10.4节）。多数降低复杂度的技术要么是减小维特比译码中幸存序列的个数，要么是减少ISI通过在维特比检测器中的预处理或者判决反馈扩展的符号个数。与逐符号的均衡相比，这些低复杂度的MLSE在性能和复杂度之间取得了较好的平衡，其性能可以接近最优的MLSE，而复杂度要低得多。

8.5节所介绍的Turbo译码思想也可以用在均衡中[14~15]，相应称为Turbo均衡。Turbo均衡用一个最大后验概率（maximum a posteriori，MAP）均衡器和一个译码器之间的

反复迭代来确定发送符号。MAP均衡器用信道输出计算出发送符号的后验概率（a posteriori probability，APP），译码器再用这个输出计算出发送符号的后验概率，这些概率构成MAP均衡器和译码器之间交互的软信息，经过一定次数的迭代后，Turbo均衡器将收敛到它对发送序列的估计值上。

365

如果发送端已知信道信息，则可进行*预均衡*（pre-equalize），即让发送符号先通过一个反转信道频率响应的滤波器。由于信道反转是在发送端而不是在接收端，所以没有噪声增强的问题。对于时变信道来说，预均衡实现起来比较困难，因为它要求发送端精确已知信道。但在相对静止的有线信道中，预均衡是一种比较实际的方法。预均衡的一个缺点是信道反转会增加发送信号的动态范围，这将引起放大失真或者会降低放大器的效率，这个问题可通过Tomlinson-Harashima预编码技术来解决[16-17]。

11.8 自适应均衡：训练和跟踪

前述的所有均衡器设计都假设已知组合信道的冲激响应$h(t)=g(t)*c(t)$。在设计接收机的时候，通常我们并不知道$c(t)$，因此所设计的均衡器必须能根据具体的$c(t)$值进行调整。无线信道的响应$c(t)=c(\tau, t)$是随时间变化的，因此系统必须要定期估计信道$c(t)$并更新相应的均衡器系数，这一过程称为均衡器训练或自适应均衡[18-19]。均衡器也可以用来检测数据来调整均衡器系数，这个过程称为均衡器跟踪。盲均衡器不用训练，它用已检测的数据得到信道响应[20-23]。

在训练期间，发送端发送已知的训练序列。在k时刻，均衡器根据训练序列$[d_{k-M},\cdots,d_k]$更新抽头系数。训练序列的长度$M+1$取决于抽头个数和训练算法的收敛速度。注意当信道变得不相关后，必须要重新训练，即至少每隔T_c秒要训练一次，T_c为信道的相干时间。如果训练算法比信道相干时间还慢，那么未等到均衡器估计出信道时，信道就已经改变了。即若$(M+1)T_s>T_c$，那么在均衡器完成训练之前，信道已经不相关了。这种情况下，均衡并不是消除ISI的有效对策，此时需采用其他的技术，如多载波调制或CDMA。

设$\left\{\hat{d}_k\right\}$是发送训练序列$\{d_k\}$时，均衡器输出的判决比特。我们根据到k时刻为止收到的训练序列来更新$k+1$时刻的N个均衡器系数$\{w_{-L}(k+1),\cdots, w_L(k+1)\}$。假设采用MMSE准则，则$\{w_{-L}(k+1),\cdots, w_L(k+1)\}$应使$d_k$和$\hat{d}_k$之间的均方误差最小。$\hat{d}_k=\sum_{i=-L}^{L}w_i(k+1)y_{k-i}$，其中$y_k=y[k]$为图11-3中$k$时刻的取样输出。能使均方误差最小化的$\{w_{-L}(k+1),\cdots, w_L(k+1)\}$可以通过维纳滤波获得[4-5]：

$$\mathbf{w}(k+1)=\{w_{-L}(k+1),\cdots,w_L(k+1)\}=\boldsymbol{R}^{-1}\boldsymbol{p} \tag{11-37}$$

366

其中$\boldsymbol{p}=d_k[y_{k+L}\cdots\ y_{k-L}]^T$，$\boldsymbol{R}$为

$$\boldsymbol{R}=\begin{bmatrix} |y_{k+L}|^2 & y_{k+L}y_{k+L-1}^* & \cdots & y_{k+L}y_{k-L}^* \\ y_{k+L-1}y_{k+L}^* & |y_{k+L-1}|^2 & \cdots & y_{k+L-1}y_{k-L}^* \\ \vdots & \ddots & \ddots & \vdots \\ y_{k-L}y_{k+L}^* & \cdots & \cdots & |y_{k-L}|^2 \end{bmatrix} \tag{11-38}$$

注意到这种情况下的最优抽头更新需要求矩阵的逆，每次迭代（每符号周期T_s）需要N^2

到N^3的乘法运算。虽然复杂度较高，但这一算法的收敛速度非常快：对于N抽头的均衡器，收敛时间大约为N个符号周期。

如果复杂度是个问题，那么需要大量乘法运算的MMSE训练就是不可行的。一种更为简单的技术是最小均方（least mean square，LMS）算法（见文献[4]的11.1.2节）。这一算法对抽头系数向量$w(k+1)$进行线性更新：

$$w(k+1)=w(k)+\Delta\varepsilon_k[y_{k+L}^*\cdots y_{k-L}^*] \tag{11-39}$$

其中$\varepsilon_k=d_k-\hat{d}_k$为判决比特和训练序列之间的误差，Δ为算法中的步长。Δ的选取涉及算法的收敛速度和稳定性。当Δ值较小时，收敛速度非常慢，收敛到正确的均衡系数需要的时间大大超过N比特时间。而如果Δ值较大时，算法有可能不稳定，因为每次迭代时可能会跳过正确的抽头值。因此，性能良好的LMS算法通常取较小的Δ值，其收敛也较慢，但复杂度要比MMSE低得多，每次迭代只需要大约$2N+1$次乘法运算，复杂度与抽头数成线性关系。其他一些算法，如RLS（root least square），平方根LS（square-root-least-squares）、快速卡尔曼等的复杂度和性能介于LMS（收敛慢但复杂度低）和MMSE（收敛快但复杂度非常高）之间，详见文献[4]的第11章。表11-1列出了这些算法的乘法运算数、收敛速度等性能。

367

均衡器输出的符号估计值$\hat{d}_k$一般要经过一个门限比较器判决为最近的星座点。在传输数据的时候，可以用判决前后的误差来调整均衡器的系数，这一过程称为跟踪。跟踪的基本思想是：(i) 如果判决误差不是零，则训练一定是不理想的；(ii) 可以用判决误差来调整均衡器中的信道估计。其工作过程是用均衡器的估计输出$\hat{d}_k$和判决输出$\hat{\hat{d}}_k$来估计等效基带组合信道$F(z)$，即用MMSE调整估计的$F(z)$的系数以使$\hat{d}_k$和$\hat{\hat{d}}_k$之间的均方误差最小，再认为这个更新后的$F(z)$等于合成信道，用它来更新相应的均衡器系数，详见文献[4~5]。表11-1同时给出了不同算法的训练和跟踪特性。快速卡尔曼判决反馈和平方根LS判决反馈在收敛和跟踪时有可能不稳定，这是它们收敛快、复杂度相对低所付出的代价。

表11-1　均衡器的训练和跟踪特性

算　法	每码元周期T_s的乘法运算数	复杂度	训练收敛时间	跟踪
LMS	$2N+1$	低	慢($>10NT_s$)	差
MMSE	N^2到N^3	非常高	快($: NT_s$)	好
RLS	$2.5N^2+4.5N$	高	快($: NT_s$)	好
快速卡尔曼DFE	$20N+5$	相对低	快($: NT_s$)	好
平方根LS DFE	$1.5N^2+6.5N$	高	快($: NT_s$)	好

例11.5：假设一个五抽头的均衡器必须每隔$0.5T_c$重新训练一次，T_c为信道相干时间。传输的数据和训练序列均为BPSK调制（$T_s=T_b$），速率为1Mbit/s。比较LMS均衡和快速卡尔曼反馈判决均衡所需要的训练序列长度。若多普勒频移为80Hz，那么定期训练将使这两种均衡器的速率分别降低多少？它们各需要多少运算量？

解：均衡器必须每隔$0.5T_c=0.5/B_d=0.5/80=6.25\text{ms}$重新训练一次。由表11-1可知，对于数据率$R_b=1/T_b=1\text{Mbit/s}$，LMS算法需要的收敛时间是$10NT_b=50\times10^{-6}\text{s}$，快速卡尔曼判决反馈均衡需要$NT_b=5\times10^{-6}\text{s}$。如果每隔6.25ms训练一次，那么LMS算法中训练所占的时间比例为$50\times10^{-6}/6.25\times10^{-3}=0.008$，因而有效数据率为$(1-0.008)R_b=0.992\text{Mbit/s}$。快速卡尔曼判决反馈均衡中，训练所占的时间比例为$5\times10^{-6}/6.25\times10^{-3}=0.0008$，有效数据率为$(1-0.0008)R_b=0.9992\text{Mbit/s}$。对于每一次训练，LMS需要约$(2N+1)\times10N=550$次的运算，而快速卡尔曼判决反馈均衡需要$(20N+5)\times N=525$次的运算，相对于今天的处理技术来说，运算量没有明显的差别。

习题

11.1　若信道的等效基带冲激响应为$H(f)=1/f$，请设计一个能完全消除ISI的连续时间均衡器$H_{eq}(f)$。假设发送信号的带宽为100kHz，信道中等效基带加性白高斯噪声的功率谱密度为N_0，求此带宽范围内均衡器输出的噪声功率，这个均衡器能否改善系统的性能？

11.2　本题研究迫零均衡中的ISI和噪声增强。考虑两个多径信道，第一个信道的冲激响应为

$$h_1(t)=\begin{cases}1 & 0\leqslant t<T_m\\ 0 & \text{其他}\end{cases}$$

第二个信道的冲激响应为

$$h_2(t)=e^{-t/T_m}\quad 0\leqslant t<\infty$$

(a) 假设发送信号是幅度为A的无限冲激序列$s(t)=\sum_{n=-\infty}^{\infty}A\delta(t-nT_b)$，脉冲间隔$T_b=T_m/2$。请计算一个比特时间$T_b$内ISI的平均功率。

368

(b) 假设$T_m=10\mu\text{s}$，发送信号BPSK调制，信道冲激响应为$h_1(t)$。若采用脉冲宽度为$T=1\mu\text{s}$的矩形脉冲成形时，求无ISI条件下信道可支持的最大数据率。如果信号带宽限制为100kHz，结果又如何？

11.3　假设信道的等效基带冲激响应为

$$h(t)=\begin{cases}e^{-t/\tau} & t\geqslant 0\\ 0 & \text{其他}\end{cases}$$

其中$\tau=6\mu\text{s}$。信道中的等效基带加性白高斯噪声的功率谱密度为N_0。

(a) 假设没有脉冲成形，也没有匹配滤波器，那么等效基带连续时间的迫零均衡器的频率响应是什么？

(b) 假设在此信道上发送的信号的频率响应为矩形，信号带宽是30kHz。求信号带宽范围内有无均衡器时的信噪比之比。提示：功率谱密度为$P(f)$的平稳随机过程的总功率为$\int P(f)\mathrm{d}f$，再经过滤波器$G(f)$后，输出过程的功率谱密度为$P(f)|G(f)|^2$。

(c) 用一个三抽头的横向滤波器来近似MMSE均衡器。可以选用任何能够合理近似时间响应的方法。

11.4　假设FIR迫零均衡器的抽头系数为$w_i=c_i$，其中$\{c_i\}$是$1/F(z)$的逆z变换。证明当时$z=e^{j\omega}$，此抽头系数能使下式最小：

$$\left|\frac{1}{F(z)}-(w_0+w_1z^{-1}+\cdots+w_Nz^{-N})\right|^2$$

11.5 某通信系统的已调信号$s(t)$的功率为10mW，载波频率为f_c，频带带宽B_s=40MHz。此信号通过了满足对称性$H(f)=H(-f)$的频率选择性衰落信道：

$$H(f)=\begin{cases}1 & f_c-20\text{ MHz}\leqslant f<f_c-10\text{ MHz}\\ 0.5 & f_c-10\text{ MHz}\leqslant f<f_c\\ 2 & f_c\leqslant f<f_c+10\text{ MHz}\\ 0.25 & f_c+10\text{ MHz}\leqslant f<f_c+20\text{ MHz}\\ 0 & \text{其他}\end{cases}$$

接收信号为$y(t)=s(t)*h(t)+n(t)$，其中$n(t)$是功率谱密度为$N_0/2=10^{-12}$ W/Hz的加性高斯白噪声。

（a）假设$y(t)$通过一个连续时间的带通迫零均衡器，求此均衡器在$f_c\pm$20MHz范围内的频率响应$H_{eq}(f)$。

369

（b）对于（a）中的均衡器，求输出信噪比。

（c）假设$s(t)$的码元周期为$T_s=1/B_s$，星座大小任意。求采用（a）中的迫零均衡后能满足$P_b<10^{-3}$的最大数据率。

11.6 假设经过ISI信道后的接收信号为

$$y(t)=\sum_{i=-\infty}^{\infty}x_if(t-iT)$$

其中$x_i=\pm1$，$f(t)$为脉冲成形、等效低通信道和接收匹配滤波器的合成的基带冲激响应。假设$f(t)=\sin(\pi t/T)/(\pi t/T)$满足无ISI的奈奎斯特准则。这一脉冲成形有两个缺点：第一，它的频谱是矩形的，难以实现。第二，脉冲尾部以$1/t$衰减，因此定时误差会导致ISI不收敛。对于本题的（b）、（c）小题，我们假设$|t|>NT$时$f(t)=0$，其中N为正整数。因为严格意义下，信号不可能时域和频域都受限，所以这个假设只是一个近似，不过对实际应用来说是一个合理的假设。

（a）证明$f(t)$的折叠谱是平坦的。

（b）假设由于定时误差，信号在$t=kT+t_0$时刻抽样，$t_0<T$。写出响应$y_k=y(kT+t_0)$的表达式，并将其分成期望项和ISI项之和。

（c）假设x_i的极性正好使ISI中的每一项都是正的（最坏情况的ISI），证明（b）中的ISI项为

$$\text{ISI}\approx\frac{2}{\pi}\sin\left(\frac{\pi t_0}{T}\right)\sum_{n=1}^{N}\frac{n}{n^2-t_0^2/T^2}$$

且当$N\to\infty$时ISI$\to\infty$。

11.7 设$g(t)=\text{sinc}(t/T_s)$，$|t|<T_s$，求对$g(t)$匹配的滤波器$g_m(t)$。并求MMSE均衡所需要的噪声白化滤波器$1/G_m^*(1/z^*)$。

11.8 证明无限冲激响应MMSE均衡的最小均方误差（11-22）可以用折叠谱$F_\Sigma(f)$表示为

$$J_{\min} = T\int_{-0.5/T}^{0.5/T} \frac{N_0}{F_\Sigma(f)+N_0}\mathrm{d}f$$

11.9　证明MMSE均衡中抽头系数的梯度为

$$\nabla_{\mathbf{w}} J = \left(\frac{\partial J}{\partial w_0}, \ldots, \frac{\partial J}{\partial w_N}\right) = 2\mathbf{w}^T \mathbf{M}_v - 2\mathbf{v}_d$$

将其置为0，可得到最优抽头系数为

$$\mathbf{w}_{\text{opt}} = (\mathbf{M}_v^T)^{-1}\mathbf{v}_d^H$$

11.10　证明式（11-27）给出的无限冲激响应MMSE均衡的最小均方误差$J_{\min}$满足$0\leqslant J_{\min}\leqslant 1$。

11.11　假设信道的离散时间等效模型是三抽头$C(z)=1+0.5z^{-1}+0.3z^{-2}$，比较MMSE均衡和判决反馈均衡的最小均方误差$J_{\min}$。

11.12　本题将展示这样两点：(i) 迫零均衡器中的噪声增强；(ii) MMSE对噪声增强的消除。假设频率选择性衰落信道的等效基带频率响应为　370

$$H(f)=\begin{cases}1 & 0\leqslant |f|<10\text{kHz}\\ 1/2 & 10\text{kHz}\leqslant |f|<20\text{kHz}\\ 1/3 & 20\text{kHz}\leqslant |f|<30\text{kHz}\\ 1/4 & 30\text{kHz}\leqslant |f|<40\text{kHz}\\ 1/5 & 40\text{kHz}\leqslant |f|<50\text{kHz}\\ 0 & \text{其他}\end{cases}$$

负频率部分是对称的。信道中的加性高斯白噪声等效至基带的功率谱密度为$N_0=10^{-9}$。

(a) 求能完全消除$H(f)$引入的ISI的模拟迫零均衡器。

(b) 求 (a) 中均衡器的输出噪声功率。

(c) 假设MMSE模拟均衡器形为$H_{eq}(f)=1/(H(f)+\alpha)$。求$\alpha=0.5$和$\alpha=1$时均衡器的输出噪声功率。

(d) 定性说明信道通过$H(f)$传输，再经过MMSE模拟均衡器$H_{eq}(f)=1/(H(f)+\alpha)$ $(\alpha>0)$时信号受到的两种影响。如何选择α?

(e) 当$\alpha\to\infty$时 (c) 中的噪声功率是多少？取$\alpha\to\infty$对均衡器设计有何不利因素？

(f) 对 (c) 中的均衡器，假设系统数据速率为100kbit/s，均衡器需要1000比特的训练序列。为了使抽头系数能在信道不相关之前收敛，可允许的最大多普勒频移为多少？

11.13　为什么在数据传输时进行信道跟踪的均衡器仍然需要定期训练？列出跟踪的两种好处。

11.14　假设某个四抽头的均衡器必须每隔$0.5T_C$重新进行训练，T_C为信道相干时间。如果一个数字处理芯片每秒能处理一千万次的乘法运算，采用LMS和RLS算法的判决反馈均衡器分别需要1000次迭代（符号周期）和50次迭代，那么当采用BPSK调制，多普勒扩展为B_D=100Hz时，两种均衡器可支持的最大数据率为多少？B_D=1000Hz呢？假设训练序列和信息序列的传输速率相同。

11.15　本题考虑均衡器跟踪时的信道估计更新。假设目标是最小化$\hat{d}_k$和$\hat{\hat{d}}_k$之间的均方误差，对于信道$H(z)$，请给出相应的信道系数更新公式。　371

11.16 超宽带（ultrawideband，UWB）系统将数据信号及其功率扩展到一个非常宽的带宽上，使得每赫兹上的信号功率非常低，通常低于噪声。因此，UWB系统可以和其他系统共存而没有较大的干扰。考虑一个BPSK调制的基带超宽带系统。数据比特采用矩形脉冲$g(t)$，$g(t)$的持续时间T远小于比特周期T_b，本题假设$T=10^{-9}$。于是BPSK调制的UWB信号为 $s(t)=\sum_n d_n g(t-nT_b)$ 其中d_n取±1，$T_b>>T$为比特周期。图11-6示出了数据序列为0、1交替时$s(t)$的波形。

（a）若$T_b=10^{-5}$，$s(t)$的带宽近似为多少？

（b）UWB信号的优点之一是在一般的信道上不会经历平衰落。考虑单比特发送：$s(t)=d_0 g(t)$，$s(t)$通过一个两径信道$h(t)=\alpha_0\delta(t)+\alpha_1\delta(t-\tau)$传输。请画出$\tau<<T$和$\tau>>T$时的信道输出。这两个图中那一个更可能是实际无线信道的输出？这个结果为什么能表明UWB信号一般不会经历平衰落？

（c）假设信道的多径时延扩展为$T_m=20\mu s$，根据图11-6，此信道实际可支持的无ISI传输的最大数据率为多少？此数据率下$s(t)$的带宽小于信道相关带宽吗？

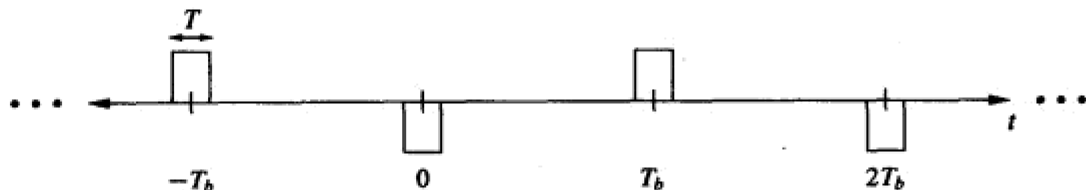


图11-6 习题11.16中UWB信号的形式

（d）假设$F(z)=\alpha_0+\alpha_1 z^{-1}+\alpha_2 z^{-2}$表示UWB系统中发送、信道和匹配滤波器合成的冲激响应。请用一个两抽头的数字均衡器$H_{eq}(z)=w_0+w_1 z^{-1}$来近似针对$F(z)$的IIR迫零均衡器，可以采用任意的近似方法。

（e）对于（d）中的均衡器，假设系统的数据速率为100kbit/s，均衡器需要1000比特的训练序列，那么为了使均衡器系数能在信道不相关之前收敛，可允许的最大多普勒频移为多少？

参考文献

[1] E. H. Satorius and S. T. Alexander, "Channel equalization using adaptive lattice algorithms," *IEEE Trans. Commun.*, pp. 899–905, June 1979.
[2] F. Ling and J. Proakis, "Adaptive lattice decision feedback equalizers – Their performance and application to time-variant multipath channels," *IEEE Trans. Commun.*, pp. 348–56, April 1985.
[3] J. Cioffi and T. Kailath, "Fast, recursive-least-squares transversal filters for adaptive filtering," *IEEE Trans. Signal Proc.*, pp. 304–37, April 1984.
[4] J. G. Proakis, *Digital Communications*, 4th ed., McGraw-Hill, New York, 2001.
[5] G. L. Stuber, *Principles of Mobile Communications*, 2nd ed., Kluwer, Dordrecht, 2001.
[6] G. D. Forney, Jr., "Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference," *IEEE Trans. Inform. Theory*, pp. 363–78, May 1972.
[7] B. Sklar, "How I learned to love the trellis," *IEEE Signal Proc. Mag.*, pp. 87–102, May 2003.
[8] R. Price, "Nonlinearly feedback-equalized PAM vs. capacity," *Proc. IEEE Internat. Conf. Commun.*, pp. 22.12–22.17, June 1972.
[9] J. Salz, "Optimum mean-square decision feedback equalization," *Bell System Tech. J.*, pp. 1341–73, October 1973.

372

[10] M. V. Eyuboglu, "Detection of coded modulation signals on linear, severely distorted channels using decision-feedback noise prediction with interleaving," *IEEE Trans. Commun.*, pp. 401–9, April 1988.

[11] J. M. Cioffi, G. P. Dudevoir, V. Eyuboglu, and G. D. Forney, Jr., "MMSE decision-feedback equalizers and coding. Part I: Equalization results," *IEEE Trans. Commun.*, pp. 2582–94, October 1995.

[12] J. M. Cioffi, G. P. Dudevoir, V. Eyuboglu, and G. D. Forney, Jr., "MMSE decision-feedback equalizers and coding. Part II: Coding results," *IEEE Trans. Commun.*, pp. 2595–2604, October 1995.

[13] J. M. Cioffi and G. D. Forney, Jr., "Generalized decision-feedback equalization for packet transmission with ISI and Gaussian noise," in A. Paulraj, V. Roychowdhury, and C. Schaper (Eds.), *Communication, Computation, Control, and Signal Processing*, pp. 79–127, Kluwer, Boston, 1997.

[14] C. Douillard, M. Jezequel, C. Berrou, A. Picart, P. Didier, and A. Glavieux, "Iterative correction of intersymbol interference: Turbo equalization," *Euro. Trans. Telecommun.*, pp. 507–11, September/October 1995.

[15] M. Tüchler, R. Koetter, and A. C. Singer, "Turbo equalization: Principles and new results," *IEEE Trans. Commun.*, pp. 754–67, May 2002.

[16] M. Tomlinson, "A new automatic equalizer emplying modulo arithmetic," *Elec. Lett.*, 7, pp. 138–9, 1971.

[17] H. Harashima and H. Miyakawa, "Matched-transmission techniques for channels with intersymbol interference," *IEEE Trans. Commun.*, pp. 774–80, August 1972.

[18] J. G. Proakis, "Adaptive equalization for TDMA digital mobile radio," *IEEE Trans. Veh. Tech.*, pp. 333–41, May 1991.

[19] S. U. Qureshi, "Adaptive equalization," *Proc. IEEE*, pp. 1349–87, September 1985.

[20] A. Benveniste and M. Goursat, "Blind equalizers," *IEEE Trans. Commun.*, pp. 871–83, August 1984.

[21] C. R. Johnson Jr., "Admissibility in blind adaptive channel equalization," *IEEE Control Syst. Mag.*, pp. 3–15, January 1991.

[22] R. Johnson, P. Schniter, T. J. Endres, J. D. Behm, D. R. Brown, and R. A. Casas, "Blind equalization using the constant modulus criterion: A review," *Proc. IEEE*, pp. 1927–50, October 1998.

[23] L. Tong, G. Zu, and T. Kailath, "Blind identification and equalization based on second-order statistics: A time domain approach," *IEEE Trans. Inform. Theory*, pp. 340–9, March 1994.

373



第12章 多载波调制

多载波调制的基本思想是将传输比特流分成多个子比特流，再调制到不同的子载波上进行传输。一般情况下，各子载波在理想传播条件下是相互正交的。子载波上的数据速率远小于总数据速率，各子信道的带宽也远小于系统总带宽。选择子载波个数时，应使每个子信道的带宽小于信道的相干带宽，使每个子信道所经历的衰落是相对平坦的，从而让子信道上的码间串扰比较小。多载波调制中的子信道不一定要相互邻近，所以高速多载波通信并不要求大段连续的频带。此外，多载波调制可以用数字方式高效率地实现。正交频分复用（orthogonal frequency division multiplexing，OFDM）这种数字实现方式通过引入循环前缀可以完全消除ISI。

许多无线通信系统已经采用了多载波调制技术，但它并不是一项全新的技术。早在20世纪50年代末至60年代初就已出现在军用高频无线通信中。大约从1990年起[1]，多载波调制技术开始应用于各种有线及无线通信中，包括欧洲的数字音频和数字视频广播[2]，采用离散多音调制的数字用户线路（digital subscriber line，DSL）[3~5]，以及新一代的无线局域网[5~7]。多载波调制还有许多新的应用，包括固定无线宽带业务[8~9]，以FLASH-OFDM著称的移动无线宽带业务[10]，甚至包括超宽带无线通信，其中多频带OFDM是IEEE 802.15超宽带标准中两个竞争的提案之一。多载波调制也是下一代蜂窝系统空中接口的候选方案之一[11~12]。

多载波技术可通过多种方式实现，包括矢量编码[13~14]和OFDM[13][15]。本章将就这些方法进行讨论。这些技术差异不大，其原理都是将宽带信道通过正交分解分割成许多并行的窄带信道。

374

对于多载波调制和单载波调制哪个更适用于码元级时延扩展的ISI信道，业内还有一些争议。文献[2]认为，在某些移动无线通信中，带均衡器的单载波系统的性能大致与包含信道编码、频域交织、加权最大似然译码的多载波调制性能相近。但文献[2]没有考虑自适应加载，它有希望显著提高多载波系统的性能[16]。不过，多载波调制也还存在一些影响其性能的问题，其中最严重的是频率偏移和定时抖动会降低子信道的正交性。此外，多载波系统的峰均功率比远大于单载波系统，这不利于使用非线性放大器。考虑到这些弊端，文献[17]讨论了多载波与单载波分组传输之间折中的问题。

虽然存在不足之处，多载波技术还是广泛用于具有较大时延扩展的高速无线通信中，因为它显著优于时域均衡。尤其在高速数据系统中，用均衡器来达到较好的性能时，需要的抽头数会很多，实现也相当复杂。而且在快速变化的信道中，让大量抽头系数都保持精确匹配也很困难。由于这些原因，大多数新出现的高速无线通信系统都不是采用均衡技术来补偿ISI，而是用多载波调制或扩频技术。

12.1 多载波数据传输

最简单的多载波调制方式是将数据流分成多个子比特流，再调制于不同中心频率的正交子载波上。选择子比特流的个数时应使每个子比特流的码元周期远大于信道的时延

PDG

扩展，子比特流的带宽相应地将远小于信道的相干带宽，这样就可以保证子比特流没有明显的ISI。

假设线性调制系统的数据传输率为R、通带带宽为B。若信道的相干带宽$B_c>B$，就会经历频率选择性衰落。多载波调制的基本原理是将这种宽带系统分割成N个并行的线性调制子系统，每个子信道的带宽为$B_N=B/N$，数据速率$R_N\approx R/N$。如果N足够大，那么子信道的带宽$B_N=B/N=B_c$，这样，每个子信道的衰落就是相对平坦的。从时域来看，子信道上的码元周期T_N与子信道带宽的倒数$1/B_N$成正比，故$B_N=B_c$意味着$T_N\approx 1/B_N\geqslant 1/B_c\approx T_m$，$T_m$为信道的时延扩展。如果$N$足够大，码元周期将远大于时延扩展，从而使各子信道上的ISI很小。

375

图12-1示出了一个多载波发射机[1]。输入的比特流通过串并变换分为N个子比特流。用QAM或PSK等线性调制方式将第n个子比特流调制到子载波f_n上，带宽为B_N。我们假设各子载波为相干解调，这样，分析中子载波的相位可以忽略。若采用滚降系数为β的升余弦脉冲$g(t)$，那么子比特流的码元周期为$T_N=(1+\beta)/B_N$。各子信道上的信号叠加成为总的发送信号：

$$s(t)=\sum_{i=0}^{N-1}s_i g(t)\cos(2\pi f_i t+\phi_i) \tag{12-1}$$

其中s_i是第i个子载波上的复数符号，ϕ_i是第i个子载波的初相。为使子信道不重叠，令$f_i=f_0+i(B_N)$, $i=0, \cdots, N-1$，这样，每个子比特流占用一个带宽为B_N的正交子信道，总带宽为$NB_N=B$，总数据传速率为$NR_N\approx R$。由此可见，这种多载波调制并没有改变原系统的数据速率或信号带宽，但因为$B_N=B_c$，它几乎没有ISI。

图12-2是这种多载波调制的接收机。每个子比特流先经过一个窄带滤波器以滤除其他子比特流，再经解调、串并变换后合为原始数据流。子信道i受平衰落的影响，其信道增益是$\sigma_i=H(f_i)$。

上面这种多载波调制虽然容易理解，但它存在一些严重的缺点。首先，在实际实现中，子信道所占的带宽要比理想升余弦脉冲成形时大一些，因为实际的脉冲必然是时间受限的。设ε/T_N是因时间受限而增加的带宽，则子信道的频率间隔必须是$(1+\beta+\varepsilon)/T_N$，由于多载波系统有$N$个子信道，那么由于时域受限总共要多花费的带宽为$\varepsilon N/T_N$。为了保持子载波正交需要的总带宽为

376

$$B=\frac{N(1+\beta+\varepsilon)}{T_N} \tag{12-2}$$

所以，这种多载波调制的频谱效率较低。另外它的接收端还需要接近理想特性的滤波器以保持子载波互不干扰，这就增加了成本。更重要的是，这种方法需要N个单独的调制解调器，这将显著增加成本、体积和功耗。下一节将介绍一种允许子载波相互重叠、不需要严格滤波的调制方法。12.4节将介绍多载波调制的数字实现方法，这种方法不需要多个调制解调器。

---

1. 实际中，复信号s_i的实部和虚部分别通过同相支路和正交支路传送。为了简化，我们在这里示出的多载波系统是在同相支路传输复信号。

PDG

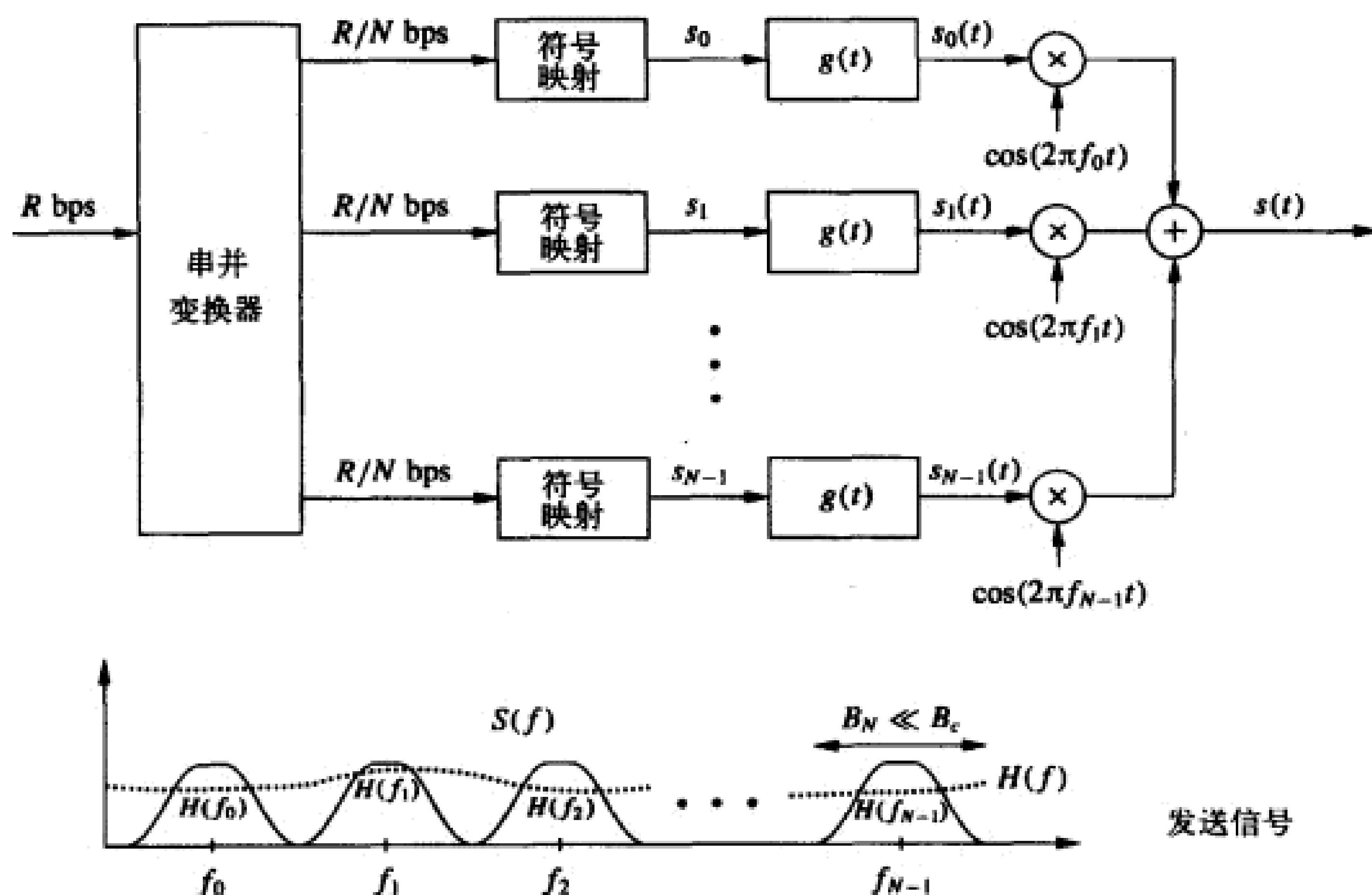


图12-1　多载波发射机

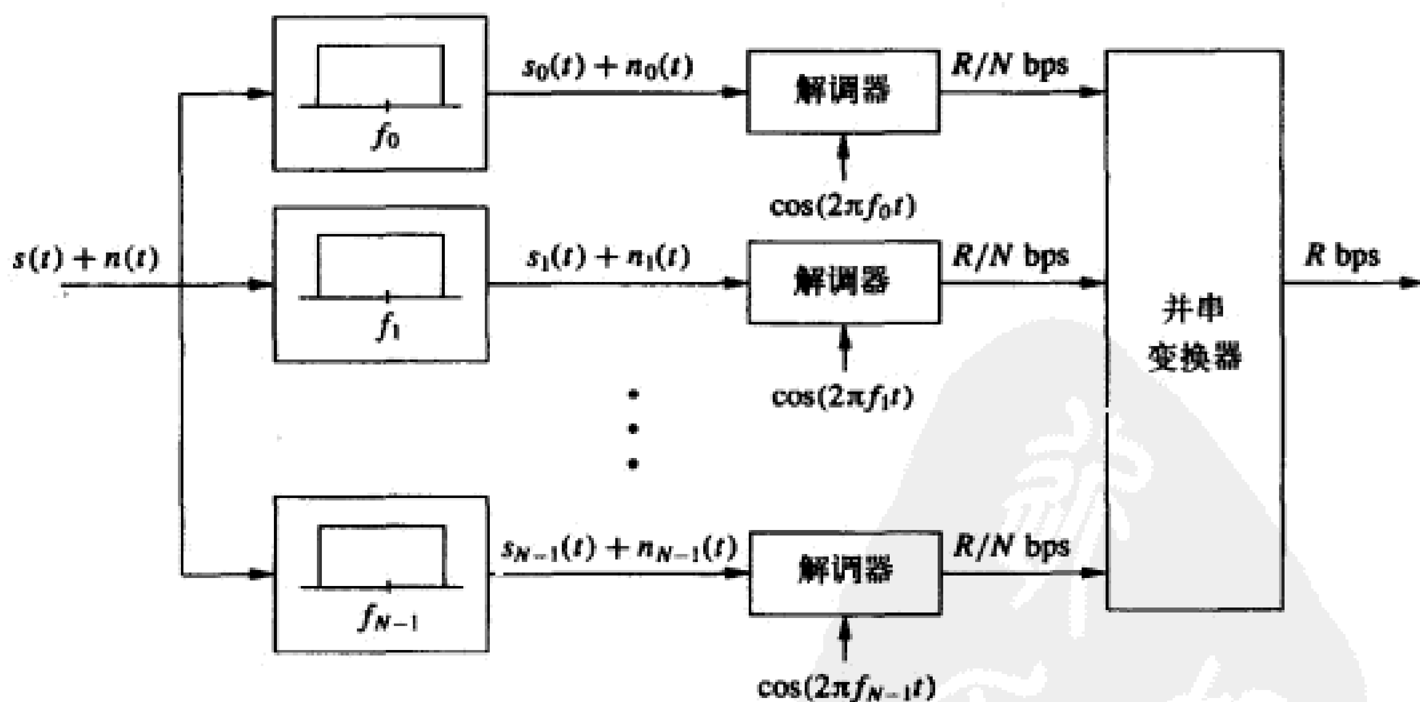


图12-2　多载波接收机

例12.1：考虑一个总通带带宽为1MHz的多载波系统，该系统在一个信道时延扩展为T_m=20μs的城市使用。为了使每个子信道都近似为平衰落，需要多少个子信道？

解：信道的相干带宽为$B_c=1/T_m=1/0.00002=50\text{kHz}$。为使每个子信道为平衰落，取$B_N=B/N=0.1B_c<<B_c$。故需要的子信道数是$N=B/0.1B_c=1000000/5000=200$。数字实现时，为了

DFT和IDFT的操作，N必须是2的整幂，此时可取N=256。 377

例12.2：一多载波系统的T_N=0.2ms: $T_N \geqslant T_m$，T_m为信道时延扩展，因此每个子信道的ISI极小。系统有128个子信道，采用滚降系数为β=1的升余弦脉冲。假设为了保证带外功率非常小，脉冲时间受限因素所需要的额外带宽为ε=0.1。请问系统的总信道带宽是多少？

解：由式（12-2），

$$B=\frac{N(1+\beta+\varepsilon)}{T_N}=\frac{128(1+1+0.1)}{0.0002}=1.344\ \text{MHz}$$

下一节中将会看到，重叠子信道可以大幅减少所需的带宽。

12.2 重叠子信道的多载波调制

通过重叠子信道可以提高多载波调制的频带利用率。为使接收端译码器可将其分离，各子信道仍必须是正交的。在时间$[0, T_N]$内，对于任意子载波初相$\{\phi_i\}$，子载波$\{\cos(2\pi(f_0+i/T_N)t+\phi_i), i=0, 1, 2, \cdots\}$近似为一个正交基，因为

$$\begin{aligned}&\frac{1}{T_N}\int_0^{T_N}\cos\left(2\pi\left(f_0+\frac{i}{T_N}\right)t+\phi_i\right)\cos\left(2\pi\left(f_0+\frac{j}{T_N}\right)t+\phi_j\right)\mathrm{d}t\\&\quad=\frac{1}{T_N}\int_0^{T_N}0.5\cos\left(2\pi\frac{(i-j)t}{T_N}+\phi_i-\phi_j\right)\mathrm{d}t\\&\quad\quad+\frac{1}{T_N}\int_0^{T_N}0.5\cos\left(2\pi\left(2f_0+\frac{i+j}{T_N}\right)t+\phi_i+\phi_j\right)\mathrm{d}t\\&\quad\approx\frac{1}{T_N}\int_0^{T_N}0.5\cos\left(2\pi\frac{(i-j)t}{T_N}+\phi_i-\phi_j\right)\mathrm{d}t\\&\quad=0.5\delta(i-j)\end{aligned}\tag{12-3}$$

近似成立是因为式中的第二项积分式在f_0T_N>>1时约等于0。还可证明，对于任意的初相，在时间$[0, T_N]$内不可能用更小的频率间隔构成正交基。即在符号间隔$[0, T_N]$上能保持子载波正交的最小频率间隔为$1/T_N$。由于载波是正交的，由5.1节可知，适当选取基带成形脉冲$g(t)$后，$\{g(t)\cos(2\pi(f_0+i/T_N)t+\phi_i), i=0, 1, 2, \cdots N-1\}$近似为一组正则基函数。一般可选$g(t)$为升余弦脉冲。下面我们将看到，采用这种正交基时，即使子信道相互重叠，接收端也能分离出各子信道上传输的调制信号。

假设多载波调制的每个子信道采用滚降系数为β的升余弦脉冲，则每个子信道的通带带宽为$B_N=(1+\beta)/T_N$。令第i个子载波的频率为(f_0+i/T_N)，$i=0, 1, \cdots, N-1$，则子载波间隔为$1/T_N$。β>0时每个子信道的带宽为$B_N=(1+\beta)/T_N>1/T_N$，所以子信道是重叠的。时域受限将使每个子信道的带宽额外增加ε/T_N。但由于子信道相互重叠，所以除了第一个和最后一个子信道以外，β和ε并不会对总带宽造成影响，如图12-3所示。重叠子信道系统的总带宽为

$$B=\frac{N+\beta+\varepsilon}{T_N}\approx\frac{N}{T_N}\tag{12-4}$$

378

PDG

当N足够大时，其中的近似式成立，因此β和ε对系统总带宽的影响可以忽略。而信道不重叠时的系统需要的总带宽是$B=N(1+\beta+\varepsilon)/T_N$。

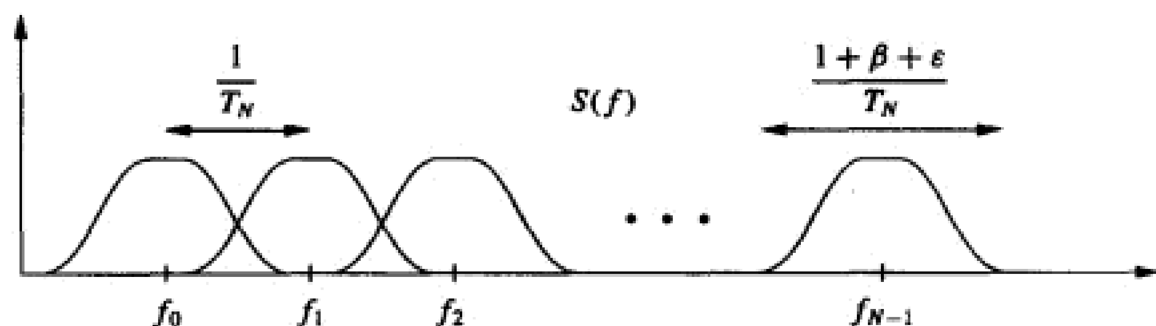


图12-3　子载波重叠的多个载波

例12.3：在例12.2的条件下，比较重叠子信道与非重叠子信道的多载波系统所需要的带宽。

解：已知$T_N=0.2\text{ms}$、$N=128$、$\beta=1$且$\varepsilon=0.1$。对于重叠子信道的系统，由式（12-4）得

$$B=\frac{N+\beta+\varepsilon}{T_N}=\frac{128+1+0.1}{0.0002}=645.5\ \text{kHz}\approx N/T_N=640\ \text{kHz}$$

而同样条件下非重叠子信道的系统需要的带宽是1.344MHz，比重叠子信道的系统两倍还多。

为了能分离出重叠的子载波，所需要的接收机结构将与图12-2不同。重叠子信道系统的接收机结构如图12-4所示，它能正确解调发送符号而不受重叠子信道的干扰。若忽略信道响应$h(t)$和噪声$n(t)$，那么对于式（12-1）给出的接收信号$s(t)$，图12-4中抽样输出为：

379

$$\begin{aligned}\hat{s}_i&=\int_0^{T_N}\left(\sum_{j=0}^{N-1}s_jg(t)\cos(2\pi f_jt+\phi_j)\right)g(t)\cos(2\pi f_it+\phi_i)\,\mathrm{d}t\\&=\sum_{j=0}^{N-1}s_j\int_0^{T_N}g^2(t)\cos\left(2\pi\left(f_0+\frac{j}{T_N}\right)t+\phi_j\right)\cos\left(2\pi\left(f_0+\frac{i}{T_N}\right)t+\phi_i\right)\mathrm{d}t\\&=\sum_{j=0}^{N-1}s_j\delta(j-i)\qquad(12\text{-}5)\\&=s_i\qquad(12\text{-}6)\end{aligned}$$

其中式（12-5）成立是因为$\{g(t)\cos(2\pi f_i t+\phi_i)\}$是$[0,T_N]$上的一组正则基。考虑信道和噪声的影响后，子信道$i$的符号有一个信道增益系数$\alpha_i=H(f_i)$，并叠加了噪声，成为$\hat{s}_i=\alpha_is_i+n_i$，其中$n_i$是功率为$N_0B_N$的白高斯噪声。

重叠子信道的多载波系统在频谱效率方面远远优于非重叠子信道的系统。但子载波相互重叠使其正交性易受定时偏差和频率偏差的影响。即使很小的偏差也会造成子信道间的相互干扰，从而大大降低系统的性能。12.5.2节将详细讨论这一问题。

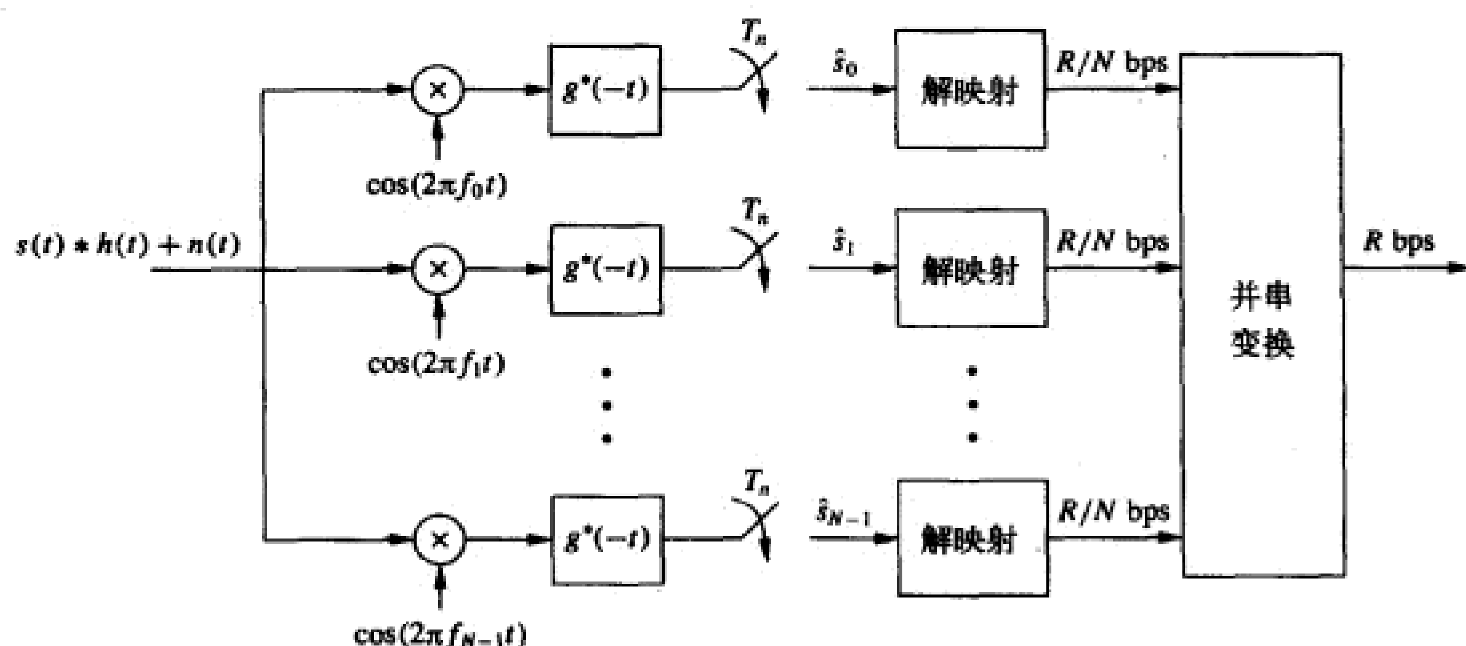


图12-4 重叠子载波系统的接收机

12.3 子载波衰落的抑制

多载波调制技术的优点是子信道带宽相对较窄，从而抑制了时延扩展的影响。但子信道上的平衰落将使某些子信道有较高的误码率。如果第i个子载波的发送功率为P_i，衰落系数为σ_i，那么接收信噪比为$\gamma_i=\alpha_i^2 P_i/(N_0 B_N)$，$B_N$是子信道带宽。若$\sigma$较小，那么第$i$个子信道的接收信噪比也很小，导致较高的误码率。在无线信道中，σ_i是服从一定衰落分布的时变函数，由它引起的性能下降与第6章所讨论的平衰落对单载波系统的影响相同。由于平衰落会严重影响子信道的传输性能，所以对子信道的衰落补偿十分重要。补偿技术有很多种，包括时频域的交织编码、频域均衡、预编码、自适应加载等，下面几节将陆续进行介绍。最常用的是交织编码，已被欧洲的数字音频及视频广播标准采纳[2,18]。在迅速变化的信道中，接收端很难估计信道特性并将其反馈至发送端。如果发送端无法得知信道特性，预编码和自适应加载就不能进行，此时只有交织编码是抑制衰落最有效的手段。 380

12.3.1 时频域交织编码

时频域交织编码先将数据比特编成码字，再将编码结果在时域和频域上进行交织，而后通过各子信道传送，从而使码字中的各个比特经历独立的衰落[19]。如果多数子信道有较高的信噪比，则接收到的码字中多数比特都是正确的，个别差的子信道上的误码会被编码纠正。这种跨子信道的编码利用了多载波系统内在的频率分集特性来纠正误码，只有整个系统带宽内有充分的频域分集时，它才能良好工作。如果信道的相干带宽较大，则子信道的衰落将高度相关，这将严重影响编码的效果。多数编码方案都假设接收端已知信道信息，信道估计一般是通过在时域和频域发送二维导频信号实现的[20]。

时频域交织编码利用了这样一点，即所有子载波上的数据都属于同一用户，可以进行联合处理。下面几节要介绍的其他抑制衰落的手段基本上都是平衰落补偿技术，这些技术除了适用于多载波系统外，也适用于窄带平衰落的单载波系统。

12.3.2 频域均衡

频域均衡基本上就是将第i个子信道的平衰落σ_i在接收端进行反转[2]，即乘以系数$1/\sigma_i$，从而使信号功率变为$\sigma_i^2 P_i/\sigma_i^2 = P_i$。这样做虽然消除了平衰落，但同时也增大了噪声。输入的噪声也乘上了系数$1/\sigma_i$，于是噪声功率变成$N_0 B_N/\sigma_i^2$，第i个子信道的信噪比在频域均衡后还是一样的。因此，频域均衡并没有真正地改善子信道平衰落引起的性能下降。

12.3.3 预编码

预编码的基本思想与频域均衡相同，只不过它是在发送端进行衰落反转[21]。这种技术要求发送端已知子信道的平衰落增益σ_i, i=0, …, $N-1$，这些增益需要通过信道估计得到[22]。若子信道i期望的接收功率为P_i，信道平衰落增益为σ_i，那么采用预编码时，子信道i的发射功率为P_i/σ_i^2。经过平衰落σ_i后，接收功率为$P_i\sigma_i^2/\sigma_i^2 = P_i$，正是所期望的功率。注意信道反转是在发送端完成而不是接收端，所以噪声功率仍然是$N_0 B_N$。预编码常用于有线多载波系统中，如高比特率数字用户线（high-bit-rate digital subscriber line，HDSL）。预编码用在无线通信中时存在两大问题。首先，预编码是基于信道反转的，从6.3.5节可知，信道反转的功率效率不高，如果是瑞利衰落信道的话，信道反转甚至需要无限大的功率。其次，预编码需要发送端精确估计信道，这对快变衰落信道来说是很困难的。

381

12.3.4 自适应加载

自适应加载技术基于第9章所讨论的自适应调制技术，常用于发送端容易得到信道估计的慢变信道中，如数字用户环路[16]。其基本思想是针对各子信道增益的不同赋予不同的数据速率和发送功率。如同预编码，自适应加载也需要发送端已知子信道的衰落$\{\sigma_i, i=0, \cdots, N-1\}$。它利用自适应调制编码技术（如可变速率、可变功率的MQAM等），通过调整各子载波的功率与速率得到最大的系统速率。

在研究自适应调制之前，我们先来考虑多载波系统的容量。假设有N个带宽为B_N的独立子信道，信道增益为$\{\sigma_i, i=0, \cdots, N-1\}$。若总功率约束为$P$，则系统的总容量为[1]：

$$C = \max_{P_i:\sum P_i = P} \sum_{i=0}^{N-1} B_N \log_2\left(1 + \frac{\alpha_i^2 P_i}{N_0 B_N}\right) \tag{12-7}$$

能使上式最大的功率分配值即为频域注水法（4-24）：

$$\frac{P_i}{P} = \begin{cases} 1/\gamma_c - 1/\gamma_i & \gamma_i \geqslant \gamma_c \\ 0 & \gamma_i < \gamma_c \end{cases} \tag{12-8}$$

1. 如4.3.1节中讨论过的，当σ_i相互独立时，这个求和式的结果正好等于系统容量。然而为了让σ_i相互独立，子信道间的间隔应不小于信道的相干带宽，这意味子信道不是平衰落。但是设计中的子信道是平衰落，故其增益$\{\sigma_i, i=1, \cdots, N\}$必是相关的，在这种情况下，各子信道的容量总和为实际系统容量的上界。我们视这个上界为实际的容量，因为实际中与这个上界相当接近。

其中$\gamma_i=\sigma_i^2 P/N_0 B_N$。将功率分配结果代入功率约束条件可解出截止值$\gamma_c$。系统容量为

$$C=\sum_{i:\gamma_i\geq\gamma_c} B_N\log_2\left(\frac{\gamma_i}{\gamma_c}\right) \tag{12-9}$$

再来考虑将9.3.2节中的可变速率与可变功率MQAM用于子信道，总数据速率为

$$R=\sum_{i=0}^{N-1} B_N\log_2\left(1+\frac{K\gamma_i P_i}{P}\right) \tag{12-10}$$

382

其中$K=-1.5/\ln(5P_b)$，P_b是子信道预期的误码率。上式对P_i优化可得到最佳功率分配为

$$\frac{KP_i}{P}=\begin{cases}1/\gamma_K-1/\gamma_i & \gamma_i\geqslant\gamma_K\\ 0 & \gamma_i<\gamma_K\end{cases} \tag{12-11}$$

相应的速率为

$$R=\sum_{i:\gamma_i\geq\gamma_K} B_N\log\left(\frac{\gamma_i}{\gamma_K}\right) \tag{12-12}$$

其中γ_k是由功率约束P和K决定的截止衰落深度。

12.4 多载波调制的数字实现

虽然20世纪50年代就已经提出了多载波调制，但它要求各子信道有单独的调制解调器，这对当时的大多数系统来说过于复杂。20年后，离散傅里叶变换（discret Fourier transform，DFT）及其逆变换（inverse DFT，IDFT）给出了简单、低成本的实现方式，同时人们认识到多载波调制可以由这些算法实现，从而使多载波调制技术获得了广泛应用。本节首先回顾DFT的基本性质，然后说明通过DFT和IDFT实现的多载波调制OFDM。

12.4.1 DFT及其性质

令$x[n]$，$0\leqslant n\leqslant N-1$，表示一个离散时间序列，其$N$点DFT定义为[23]：

$$\text{DFT}\{x[n]\}=X[i]\triangleq\frac{1}{\sqrt{N}}\sum_{n=0}^{N-1}x[n]\mathrm{e}^{-\mathrm{j}2\pi ni/N}\quad 0\leqslant i\leqslant N-1 \tag{12-13}$$

DFT等价于连续时间的傅里叶变换，$X[i]$反映原始信号$x(t)$经过时间采样后的$x(n)$的频域信息。连续时间傅里叶变换和DFT都基于这样一点，即复指数函数是线性系统的特征函数。用IDFT可从序列$x(n)$的DFT中复原出$x(n)$：

$$\text{IDFT}\{X[i]\}=x[n]\triangleq\frac{1}{\sqrt{N}}\sum_{i=0}^{N-1}X[i]\mathrm{e}^{\mathrm{j}2\pi ni/N}\quad 0\leqslant n\leqslant N-1 \tag{12-14}$$

DFT及其逆变换一般通过快速傅里叶变换（fast Fourier transfor，FFT）及其逆变换

(inverse FFT，IFFT) 硬件实现。

当输入数据流$x[n]$通过一个线性时不变的离散时间信道$h[n]$时，其输出$y[n]$为输入和信道冲激响应的离散时间卷积：

$$y[n]=h[n]*x[n]=x[n]*h[n]=\sum_k h[k]x[n-k] \tag{12-15}$$

$x[n]$与$h[n]$的N点**循环卷积**（circular convolution）定义为

383

$$y[n]=x[n]\circledast h[n]=h[n]\circledast x[n]=\sum_k h[k]x[n-k]_N \tag{12-16}$$

其中$[n-k]_N$表示$[n-k]$模N，即$x[n-k]_N$是$x[n-k]$的周期延拓。容易验证，式（12-16）中的$y[n]$的周期也是N。根据DFT的定义，时域循环卷积对应频域乘积：

$$\mathrm{DFT}\{y[n]=x[n]\circledast h[n]\}=X[i]H[i] \quad 0\leqslant i\leqslant N-1 \tag{12-17}$$

其中$H[i]$是$\{h[n]\}$的N点DFT。注意，如果序列$\{h[n]\}$的长度$k<N$，做N点DFT时需在其后补上$N-k$个零。由（12-17），如果信道与输入信号是循环卷积的关系，那么只要接收端已知$h[n]$，就可以通过对$Y[i]/H[i]$，$0\leqslant i\leqslant N-1$，进行IDFT就能复原$x[n]$。然而，信道输出是线性卷积而非循环卷积。不过这个线性卷积可以转化为循环卷积，方法如下一节所述，是在输入信号的前面添加上特定的前缀，称为**循环前缀**（cyclic prefix）。

12.4.2 循环前缀

假设信道输入序列为$x[n]=x[0],\cdots,x[N-1]$，长度为N，离散时间信道的有限冲激响应（finite impulse response，FIR）为$h[n]=h[0],\cdots,h[\mu]$，长度为$\mu+1=T_m/T_s$，T_m为信道的时延扩展，T_s是离散时间序列的抽样时间。$x[n]$的循环前缀定义为$\{x[N-\mu],\cdots,x[N-1]\}$，它是$x[n]$的后$\mu$个值。对于长为$N$的输入序列，将后$\mu$个样值增加到序列的开头，形成一个长度为$N+\mu$的新序列$\tilde{x}[n]$，$-\mu\leqslant n\leqslant N-1$，其中$\tilde{x}[-\mu],\cdots,\tilde{x}[N-1]=x[N-\mu],\cdots,x[N-1]$ $x[0],\cdots,x[N-1]$，如图12-5所示。注意在此定义下，有$\tilde{x}[n]=x[n]_N$，也即$\tilde{x}[n-k]=x[n-k]_N$，$-\mu\leqslant n-k\leqslant N-1$。

384

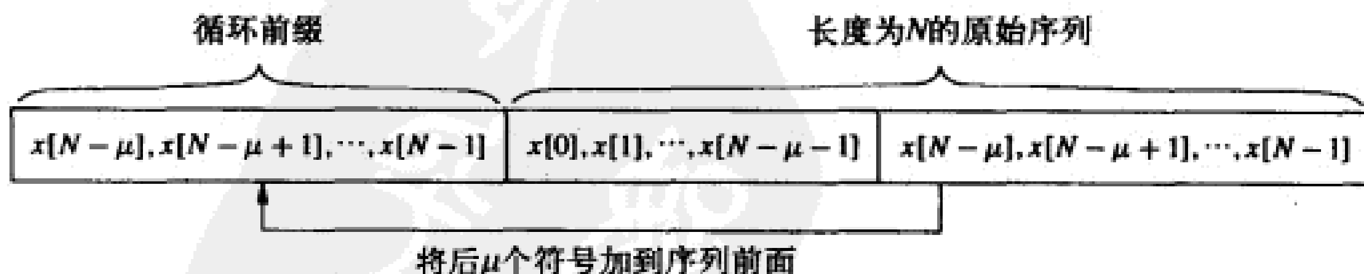


图12-5 长度为μ的循环前缀

现将$\tilde{x}[n]$输入到冲激响应为$h[n]$的离散时间信道，则输出$y[n]$, $0\leqslant n\leqslant N-1$为

$$
\begin{aligned}
y[n] &= \tilde{x}[n] * h[n] \\
&= \sum_{k=0}^{\mu} h[k]\tilde{x}[n-k] \\
&= \sum_{k=0}^{\mu} h[k]x[n-k]_N \\
&= x[n] \circledast h[n]
\end{aligned}
\tag{12-18}
$$

其中第三个等式是因为$0 \leqslant k \leqslant \mu$时 $\tilde{x}[n-k]=x[n-k]_N$，$0 \leqslant n \leqslant N-1$。因此，通过在信道输入端添加循环前缀，涉及$0 \leqslant n \leqslant N-1$时间内信道输出$y[n]$的线性卷积就转化成了循环卷积。不考虑噪声时，对信道输出作DFT得到

$$
Y[i] = \mathrm{DFT}\{y[n] = x[n] \circledast h[n]\} = X[i]H[i] \quad 0 \leqslant i \leqslant N-1 \tag{12-19}
$$

若$h[n]$已知，则可通过信道输出$y[n]$，$0 \leqslant n \leqslant N-1$恢复出输入序列$x[n]$，$0 \leqslant n \leqslant N-1$：

$$
x[n] = \mathrm{IDFT}\left\{\frac{Y[i]}{H[i]}\right\} = \mathrm{IDFT}\left\{\frac{\mathrm{DFT}\{y[n]\}}{\mathrm{DFT}\{h[n]\}}\right\} \tag{12-20}
$$

注意$y[n]$，$-\mu \leqslant n \leqslant N-1$的长度为$N+\mu$，但在式（12-20）中恢复$x[n]$，$0 \leqslant n \leqslant N-1$时不需要用到$y[n]$的前$\mu$个样值$[-\mu]$，$\cdots y[-1]$，因为循环前缀中是冗余信息。此外，若将输入序列分为一系列长为N的数据块，每个数据块都加循环前缀，那么$y[n]=h[n]* \tilde{x}[n]$的前μ个样值将受到来自前一数据块中后μ个样值的ISI的干扰，如图12-6所示。丢弃受ISI影响的信道输出的前μ个样值不会损失信息，因此循环前缀起到了有效消除数据块之间的ISI的作用。从时间上看，等于是每经过N个符号的持续时间NT，后插入了一段持续时间为T_m（信道时延扩展）的保护间隔，以消除数据块间的ISI。

循环前缀也是有代价的。在输入数据块之前加上μ个符号的前缀，开销是μ/N，导致有效数据速率降为$N/(\mu+N)$。发送循环前缀所用的功率也属浪费，因为前缀只是冗余信息。从图12-6可以看出，不管所加的长度为μ的前缀是什么，只要丢弃数据块的前μ个样值，都可以消除数据块间的ISI。特别当前缀为全零时，虽然有效数据率还是$N/(\mu+N)$，但不用花费功率来发送前缀。全零前缀是向量编码的一种，12.4.5节将讨论循环前缀和全零前缀的折中问题。

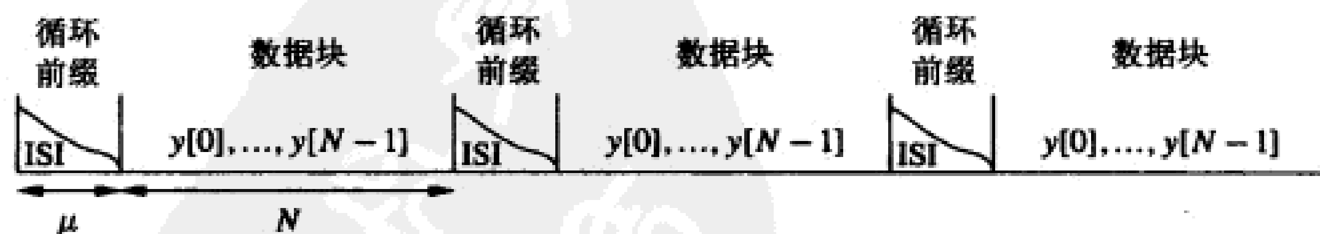


图12-6 信道输出端数据块之间的ISI

以上的这种分析推动了OFDM的设计。OFDM将输入数据分成长度为N的数据块，这些数据块叫做OFDM符号。每个OFDM符号都加循环前缀，以使输入信号与信道冲激响应进行循环卷积。接收端去除受OFDM符号间ISI影响的输出样值，用其余样值的

385 DFT变换恢复输入序列。下一节将详细地讨论OFDM系统的设计。

例12.4：设一个OFDM系统的总带宽为B=1MHz，$\beta=\varepsilon=0$。相应的单载波系统的码元间隔是$T_s=1/B=1\mu s$。信道的最大时延扩展为$T_m=5\mu s$，对单载波系统明显有严重的ISI。假设OFDM系统的每个子信道都采用16QAM调制。为使维持较低的开销，OFDM系统采用了128个子载波来消除ISI，因此$T_N=NT_s=128\mu s$。为保证OFDM符号间没有ISI，设循环前缀的长度为$\mu=8>T_m/T_s$。在这些条件下，求每个子信道的带宽、每个OFDM符号的传输时间、循环前缀的开销、以及系统的数据传输速率。

解：子信道带宽为$B_N=1/T_N=7.812$kHz，$B_N=B_c=1/T_m=200$kHz，故ISI可以忽略不计。每个OFDM符号的传输时间为$T=T_N+\mu T_s=128+8=136\mu s$。循环前缀的开销为8/128，约等于6.25%。每T秒每子载波传输了$\log_2 16=4$比特，所以数据速率为$128\times 4/136\times 10^{-6}=3.76$Mbit/s，引入循环前缀使此值略小于4Mbit/s。

12.4.3　正交频分复用（OFDM）

以OFDM方式实现的多载波调制如图12-7所示。输入数据经过QAM调制得到一个复数的数据序列$X[0], X[1], \cdots, X[N-1]$。经过串并变换后的$N$个并行QAM符号$X[0],\cdots, X[N-1]$对应$N$个子载波上传输的符号，也就是OFDM调制器输出信号$s(t)$的离散频率分量。这N个频率分量经过逆DFT变换后成为$s(t)$的时域抽样值，逆DFT可利用IFFT高效实现。IFFT所产生OFDM符号由长度为N的序列$x[n]=x[0], x[1],\cdots, x[N-1]$组成，其中

$$x[n]=\frac{1}{\sqrt{N}}\sum_{i=0}^{N-1}X[i]e^{j2\pi ni/N}\qquad 0\leqslant n\leqslant N-1\tag{12-21}$$

这个序列对应于多载波调制信号的抽样：式（12-21）右边就是用QAM符号$X[i]$对各个载波$e^{j2\pi it/T_N}$进行调制，再对这些已调信号的和进行抽样。再加入循环前缀，形成时域样值序列 $\tilde{x}[n]=x[-\mu],\cdots,x[N-1]=x[N-\mu],\cdots x[0],\cdots,x[N-1]$，经并串变换后按顺序通过

386 D/A变换器，得到OFDM基带信号$\tilde{x}(t)$，再上变频到频率f_0。

发送信号经信道冲激响应滤波后叠加了噪声，形成接收信号$r(t)$。接收信号经下变频至基带，再通过滤波器滤除高频成分，再通过A/D变换器得到样值序列$y[n]=\tilde{x}[n]*h[n]+v[n]$，$-\mu\leqslant n\leqslant N-1$，其中$h[n]$是信道的离散时间等效低通冲激响应。再去除$y[n]$的前$\mu$个样值组成的前缀。所得到的$N$个样值在无噪声时的DFT为$Y[i]=H[i]X[i]$。这些时间样值再经过串并变换及FFT后得到原始符号$H[i]\ X[i]$，其中$H[i]=H(f_i)$是第i个子信道的

387 平衰落信道增益。FFT的输出经并串变换后由QAM解调恢复出原始数据。

OFDM系统将宽带信道分解为相互正交的一组窄带子信道，每个子信道上传输不同的QAM符号。这种分解不需要已知信道增益$H[i]$, $i=0, \cdots, N-1$，如同我们把频率响应为$H(f)$的连续时间信道分割为正交信道时也不需要知道$H(f)$，只需把发送信号的频谱分割为不相重叠的子带。解调器将$Y[i]$除以信道增益$H[i]$就可以复原QAM符号，这个过程

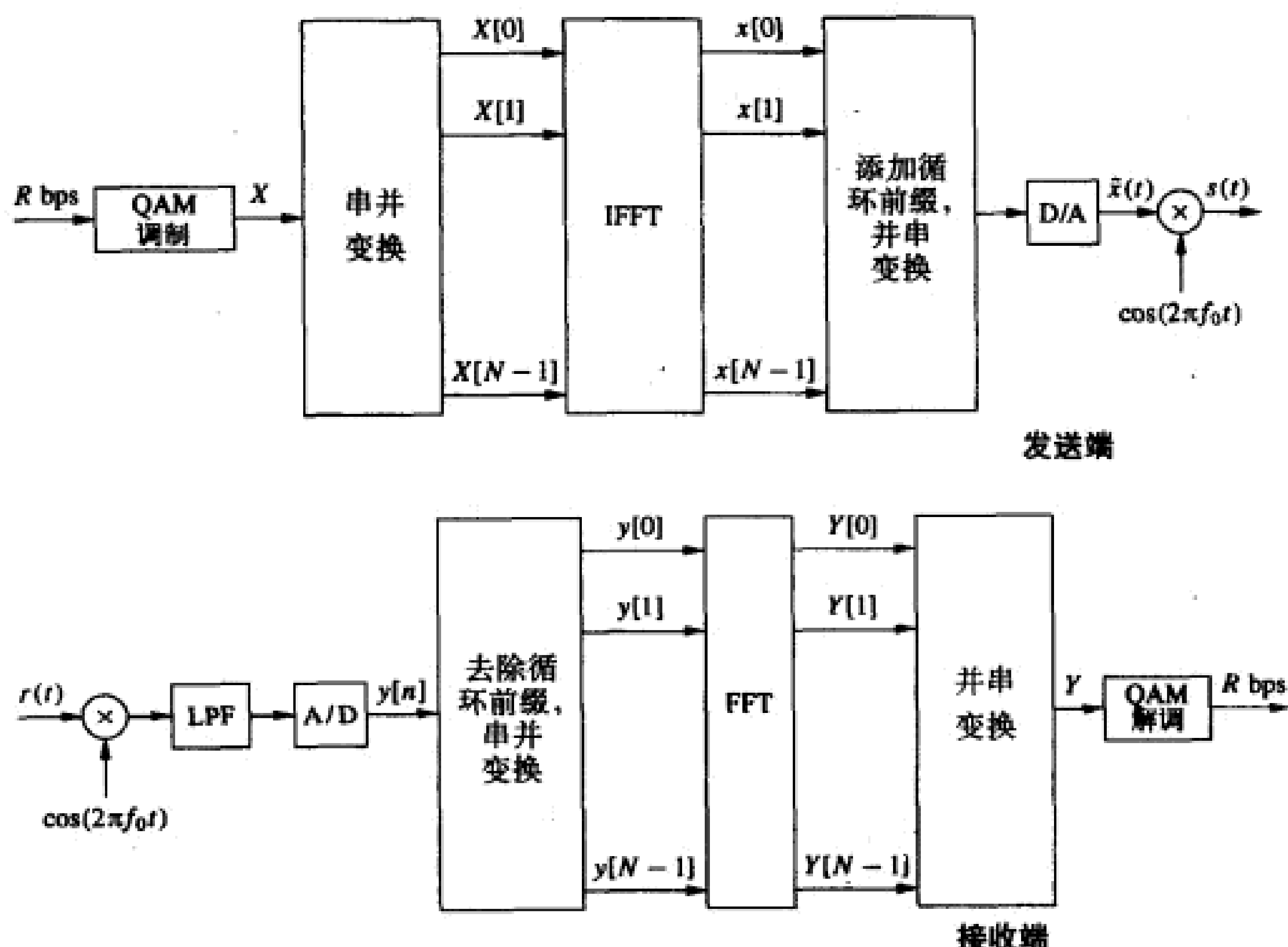


图12-7 用IFFT/FFT实现OFDM

也叫频域均衡。不过，我们在12.3.2节谈到，对于连续时间的OFDM系统，频域均衡会使噪声增强，因为第i子信道的噪声也乘上了$1/H[i]$。均衡虽然消除了平衰落对$X[i]$的影响，但接收信噪比没有改变。12.3节已经指出，克服子信道上平衰落更好的方法是预编码、自适应加载、以及跨子信道间的编码。也可以用全零前缀代替循环前缀，它在OFDM符号$x[n]$之前插入μ个全为0的样值，如图12-8所示。接收端将ISI引起的OFDM符号末端的拖尾加回到符号的首部，等效于使用了循环前缀。全零前缀不需要发送功率，相对于循环前缀，它使总发送功率降低了$N/(N+\mu)$。不过把拖尾加回首部时也加上了噪声，这使总的噪声功率加强了$(N+\mu)/N$。因此，这两种前缀的信噪比没有明显区别。

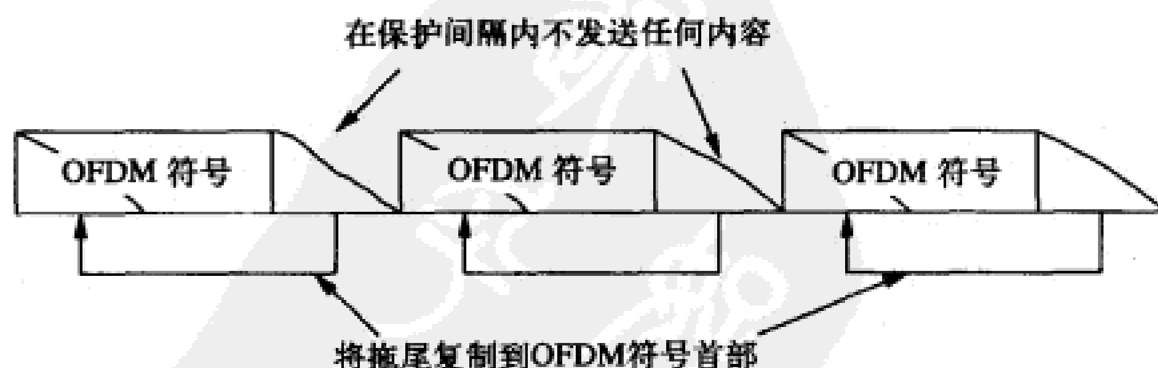


图12-8 用全零前缀产生循环信道

12.4.4 OFDM的矩阵表示

也可以用矩阵表示的方法来分析OFDM系统。考虑一个离散时间等效基带信道，其有限冲激响应为$h[n]$，$0\leqslant n\leqslant\mu$，输入为$\tilde{x}[n]$，噪声为$v[n]$，输出$y[n]=\tilde{x}[n]*h[n]+v[n]$。将这些序列中的第$n$个元素分别记为$h_n=h[n]$，$\tilde{x}_n=\tilde{x}[n]$，$v_n=v[n]$以及$y_n=y[n]$。于是输出序列可以写成如下矩阵形式

$$\begin{bmatrix} y_{N-1} \\ y_{N-2} \\ \vdots \\ y_0 \end{bmatrix} = \begin{bmatrix} h_0 & h_1 & \cdots & h_\mu & 0 & \cdots & 0 \\ 0 & h_0 & \cdots & h_{\mu-1} & h_\mu & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & h_0 & \cdots & h_{\mu-1} & h_\mu \end{bmatrix} \begin{bmatrix} x_{N-1} \\ \vdots \\ x_0 \\ x_{-1} \\ \vdots \\ x_{-\mu} \end{bmatrix} + \begin{bmatrix} v_{N-1} \\ v_{N-2} \\ \vdots \\ v_0 \end{bmatrix} \tag{12-22}$$

也可更紧凑地表示为

$$y = Hx + v \tag{12-23}$$

接收符号$y_{-1},\cdots,y_{-\mu}$受到了前一数据块的ISI影响，而且恢复数据也不需要它们，故在接收端被丢弃。$x[n]$的后μ个符号是循环前缀：$x_{-1}=x_{N-1}$、$x_{-2}=x_{N-2}$、$\cdots$、$x_{-\mu}=x_{N-\mu}$。因此，矩阵表达式（12-22）等价于

[388]

$$\begin{bmatrix} y_{N-1} \\ y_{N-2} \\ \vdots \\ \vdots \\ \vdots \\ y_0 \end{bmatrix} = \begin{bmatrix} h_0 & h_1 & \cdots & h_\mu & 0 & \cdots & 0 \\ 0 & h_0 & \cdots & h_{\mu-1} & h_\mu & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & h_0 & \cdots & h_{\mu-1} & h_\mu \\ \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ h_2 & h_3 & \cdots & h_\mu & \cdots & h_0 & h_1 \\ h_1 & h_2 & \cdots & h_{\mu-1} & \cdots & 0 & h_0 \end{bmatrix} \begin{bmatrix} x_{N-1} \\ x_{N-2} \\ \vdots \\ \vdots \\ \vdots \\ x_0 \end{bmatrix} + \begin{bmatrix} v_{N-1} \\ v_{N-2} \\ \vdots \\ \vdots \\ \vdots \\ v_0 \end{bmatrix} \tag{12-24}$$

也可更紧凑地表示为

$$y = \tilde{H}x + v \tag{12-25}$$

这个等价模型说明，插入循环前缀后，信道可以建模为N个有用样值的循环卷积矩阵$\tilde{H}$。$\tilde{H}$是正规矩阵（$\tilde{H}^H\tilde{H}=\tilde{H}\tilde{H}^H$），故可作特征值分解

$$\tilde{H} = M\Lambda M^H \tag{12-26}$$

其中Λ是由$\tilde{H}$的特征值构成的对角阵，M是由$\tilde{H}$的各个特征向量构成的酉阵。

对$x[n]$的DFT操作也可以表示为矩阵乘法

$$X = Qx$$

其中$X=(X[0],\cdots,X[N-1])^T$，$x=(x[0],\cdots,x[N-1])^T$，Q为下面给出的$N\times N$矩阵

$$Q=\frac{1}{\sqrt{N}}\begin{bmatrix}1 & 1 & 1 & \cdots & 1\\ 1 & W_N & W_N^2 & \cdots & W_N^{N-1}\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 1 & W_N^{N-1} & W_N^{2(N-1)} & \cdots & W_N^{(N-1)^2}\end{bmatrix} \tag{12-27}$$

其中$W_N=e^{-j2\pi/N}$。由于

$$Q^{-1}=Q^H \tag{12-28}$$

所以，IDFT也可类似表示为

$$x=Q^{-1}X=Q^HX \tag{12-29}$$

令v表示$\tilde{H}$ 的特征值λ的特征向量，则

$$\lambda v=\tilde{H}v$$

酉阵M的列是$\tilde{H}$ 的特征向量，即 $\lambda_i m_i=\tilde{H}m_i, i=0,1,K,N-1$，$m_i$ 表示M的第i列。用数学归纳法可以证明，DFT矩阵Q^H的列是$\tilde{H}$的特征向量，因此$Q=M^H$或$Q^H=M$。于是可以得出

$$\begin{aligned}Y&=Qy\\ &=Q[\tilde{H}x+v]\\ &=Q[\hat{H}Q^HX+v]\\ &=Q[M\Lambda M^HQ^HX+v] \qquad (12\text{-}30)\\ &=QM\Lambda M^HQ_X^H+Qv\\ &=M^HM\Lambda M^HMX+Qv\\ &=\Lambda X+v_Q \qquad (12\text{-}31)\end{aligned}$$

由于Q是幺阵，所以$v_Q=Qv$与v有相同的自相关矩阵，所以仍然是同功率的高斯白噪声。上述矩阵分析说明，通过IDFT/DFT和循环前缀和，OFDM将一个ISI信道分解成N个正交子信道，并且这样的分解不需要已知信道矩阵H。

389

在分析多天线OFDM时，这种矩阵表达很有用[24]。第10章指出，MIMO信道可用一个$M_r\times M_t$ 的矩阵表示，其中M_t 是发送天线数，M_r是接收天线数。因此，若MIMO-OFDM信道有N个子信道、M_t 个发送天线、M_r个接收天线，且信道的有限冲激响应持续时长为μ，则可以表示为

$$y=HX+v \tag{12-32}$$

其中y是$M_rN\times 1$的向量，对应M_r个接收天线上的N个时间抽样。H是$NM_r\times(N+\mu)M_t$矩阵，对应各个发送—接收天线对之间的N个平衰落子信道增益。x是$M_t(N+\mu)\times 1$向量，对应M_t 个发送天线上的N个时间样值以及长度为μ的循环前缀。v是$M_r N\times 1$的信道噪声向量。这个矩阵与单天线OFDM系统的形式相同，此前的设计与分析也同样适用：MIMO-OFDM系统把宽带信道划分为许多窄带子信道，从而消除了ISI。每个子信道都是平衰落，可按平衰落的MIMO信道处理。把标准MIMO的矩阵分析扩展到MIMO-OFDM信道可以得到信道容量[25]。数字实现时，每个发送天线的输入被分为长度为N

的数据块，添加循环前缀将线性卷积转化为循环卷积，并消除了输入数据块之间的ISI，详见文献[26]。

12.4.5　向量编码

OFDM用特征值和特征向量将$N \times N$的循环卷积信道矩阵$\tilde{H}$ 进行分解。向量编码（vector coding，VC）用类似的方法，利用奇异值分解将式（12-23）中的$N \times (N+\mu)$原始信道矩阵$\boldsymbol{H}$进行分解，这种方法适用于任意维的矩阵。奇异值分解不需要通过循环前缀使子信道正交化，因此它在能量意义下比OFDM效率更高。但它的实现也更复杂，而且矩阵分解需要已知信道的冲激响应，而OFDM的分解不需要信道信息。

由附录C，$\boldsymbol{H}$的奇异值分解可以表示为

$$\boldsymbol{H}=\boldsymbol{U}\Sigma\boldsymbol{V}^H \tag{12-33}$$

其中$\boldsymbol{U}$是$N \times N$的酉阵，$\boldsymbol{V}$是$(N+\mu) \times (N+\mu)$的酉阵，Σ的对角线上的第i个元素σ_i 等于$\boldsymbol{H}$的第i个奇异值，其余为零。$\boldsymbol{H}$的奇异值σ_i 与$\boldsymbol{HH}^H$的特征值λ_i 的关系是 $\sigma_i=\sqrt{\lambda_i}$ 。$\boldsymbol{H}$是块对角卷积矩阵，故其秩为N，即$\sigma_i \neq 0, i=1, \cdots, N$。

390

和OFDM一样，向量编码也将输入数据符号分组为有N个符号的向量。用$X_i, i=0, \cdots N-1$表示第i个子信道上的发送符号。N个发送符号和μ个后缀符号$X_N, \cdots, X_{N+\mu-1}$（一般是零）构成向量$\boldsymbol{X}=(X_0, \cdots, X_{N-1}, X_N, \cdots, x_{N+\mu-1})$。如图12-9所示，发送端用数据符号$X_i$ 乘以矩阵$\boldsymbol{V}$的第i列得到一个向量，再将这些向量相加；接收端用收到的向量y乘以$\boldsymbol{U}^H$ 的各行，得到N个输出符号Y_i, $i=0, \cdots, N-1$。这里的与V和$\boldsymbol{U}^H$ 相乘类似于MIMO中的发送预编码和接收成形。

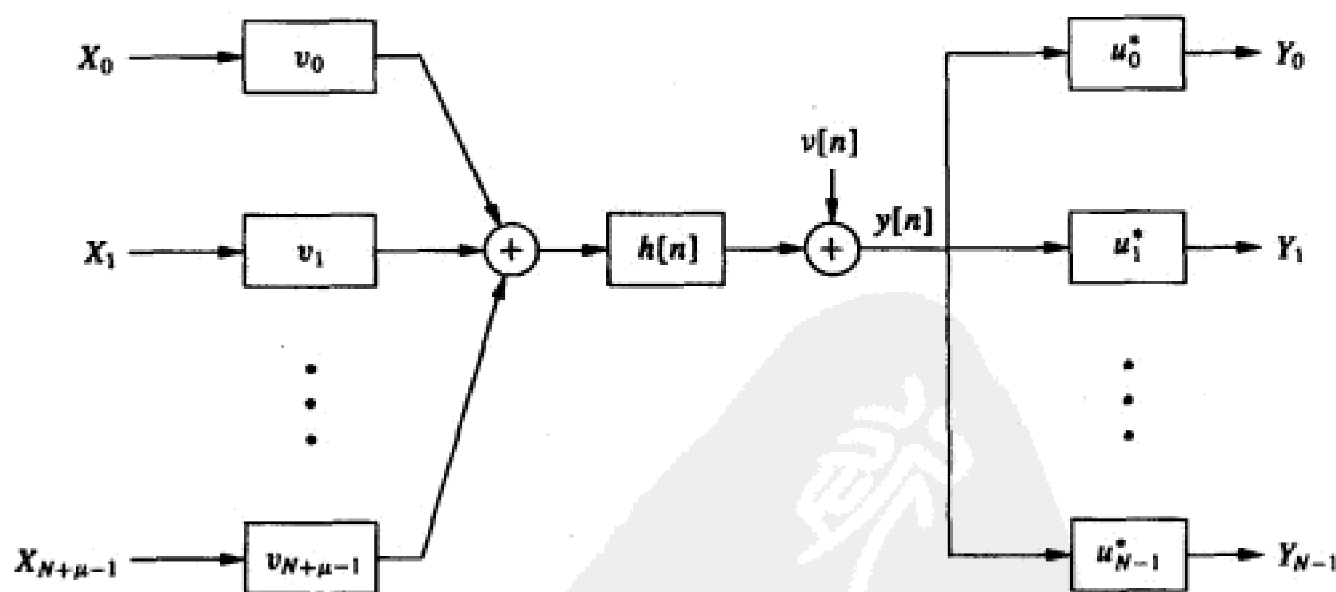


图12-9　向量编码

从数学上看，滤波后的发送和接收向量为

$$\boldsymbol{x}=\boldsymbol{VX}$$

$$\boldsymbol{Y}=\boldsymbol{U}^H\boldsymbol{y} \tag{12-34}$$

滤波后的接收向量$\boldsymbol{Y}$没有ISI，这一点用简单的线性代数即可证明：

$$
\begin{aligned}
Y &= U^H y \\
&= U^H(Hx+v) \\
&= U^H(U\Sigma V^H)VX + U^H v \\
&= \Sigma X + U^H v
\end{aligned}
\tag{12-35}
$$

因此，X的每个元素都通过了一个无ISI的标量信道，第i个子信道的标量增益是H的第个i奇异值。另外由于U是酉阵，新的噪声向量 $\tilde{v}=U^H v$ 方差不变。这样，接收向量为

$$
\begin{bmatrix} Y_{N-1} \\ Y_{N-2} \\ \vdots \\ Y_0 \end{bmatrix} = \begin{bmatrix} \sigma_1 X_{N-1} \\ \sigma_2 X_{N-2} \\ \vdots \\ \sigma_N X_0 \end{bmatrix} + \begin{bmatrix} \tilde{v}_{N-1} \\ \tilde{v}_{N-2} \\ \vdots \\ \tilde{v}_0 \end{bmatrix}
\tag{12-36}
$$

391

我们称X为向量码字（vector codeword），它是在长度为N的数据块后添加了μ个附加符号。与OFDM不同的是，由式（12-35）以及Σ的结构可以看出，奇异值分解不需要后缀的这些符号有特定的形式，不一定必须是循环前缀。当前缀为全零时，也不必把拖尾加回去。在实际中，为减小发射功率，这些附加的符号都设为零。这样一来，向量码字之间便出现了一个保护间隔或全零后缀，如图12-10所示。

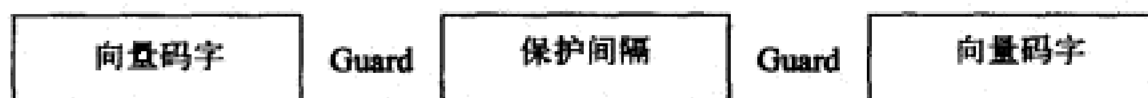


图12-10 向量编码中的保护间隔（全零后缀）

信息论及估计理论已经证明，向量编码是对N维信道H的最佳划分（见文献[13]的第4章）。所以，向量编码的容量是其他任何信道划分方法所得容量的上界。尽管它在理论上最佳，尽管它能以较小的开销得到无ISI信道，并且不浪费发射功率，向量编码在实用上还存在一些问题，其中最重要的两个问题是：

1. 复杂度：简单的多载波调制系统中，复杂度随载波数迅速增大，向量编码也如此。从图12-9可以看到，实现向量编码需要N个发送预编码滤波以及N个接收成形滤波。此外，对$N\times(N+\mu)$的矩阵H进行奇异值分解的复杂度也随N迅速增大。

2. 奇异值分解及信道信息：要将信道进行正交分解，就需要对信道矩阵H进行奇异值分解，而且发送端必须要知道预编码滤波矩阵。这就是说，每当信道发生变化时，就需要重新计算奇异值分解，并将结果通知发送端。无线通信系统一般难以承受奇异值分解的计算复杂度以及将信息通知给发送端所需要的时延。而OFDM中的分解不需要信道信息所以无线通信中的离散多载波调制选择了OFDM。

例12.5：考虑一个简单的两抽头离散时间信道（μ=1）：

$$H(z)=1+0.9z^{-1}$$

因为$\mu+1=T_m/T_s=2$，所以N=8时$B_N\approx 1/(NT_s) \ll B_c \approx 1/T_m$。试写出（12-23）形式的系统矩阵表示，并给出相应的信道矩阵H的奇异值。

392

解：对于$H(z)=1+0.9z^{-1}$及$N=8$，(12-23) 可以写为

$$\begin{bmatrix} y_7 \\ y_6 \\ \vdots \\ y_0 \end{bmatrix} = \begin{bmatrix} 1 & 0.9 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 1 & 0.9 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & 0 & 1 & 0.9 \end{bmatrix} \begin{bmatrix} x_7 \\ x_6 \\ \vdots \\ x_{-1} \end{bmatrix} + \begin{bmatrix} \nu_7 \\ \nu_6 \\ \vdots \\ \nu_0 \end{bmatrix} \tag{12-37}$$

式中矩阵$\mathbf{H}$的奇异值可以通过标准的计算机软件包（如Matlab）得到：

$$\sum=\mathrm{diag}[1.87, 1.78, 1.65, 1.46, 1.22, 0.95, 0.66, 0.34]$$

同时容易求得预编码矩阵$\mathbf{U}$和成形矩阵$\mathbf{V}$。给定$\mathbf{U}$、$\mathbf{V}$、Σ后，这个通信没有码间干扰，发送符号$X_0, \cdots, X_{L-1}$只是分别乘上了相应的奇异值，如式 (12-36) 所示。

12.5　多载波系统中的挑战

12.5.1　峰均比

峰均比（peak-to-average power ratio，PAR）是通信系统中的一个重要指标。低峰均比可以使功放高效工作，而当峰均比较高时，功放必须要有较大的回退才能保证信号的线性放大。图12-11显示出了典型的功放特性。一般要求功放工作在线性区以保证信号不失真，所以信号峰值也必须要限制在线性区内。为了使功放达到最大的效率，自然希望峰值尽量接近均值。此外，高峰均比信号的动态范围很大，这要求接收端有高分辨率的A/D变换器，相应将提高接收机前端的实现复杂度和功耗。

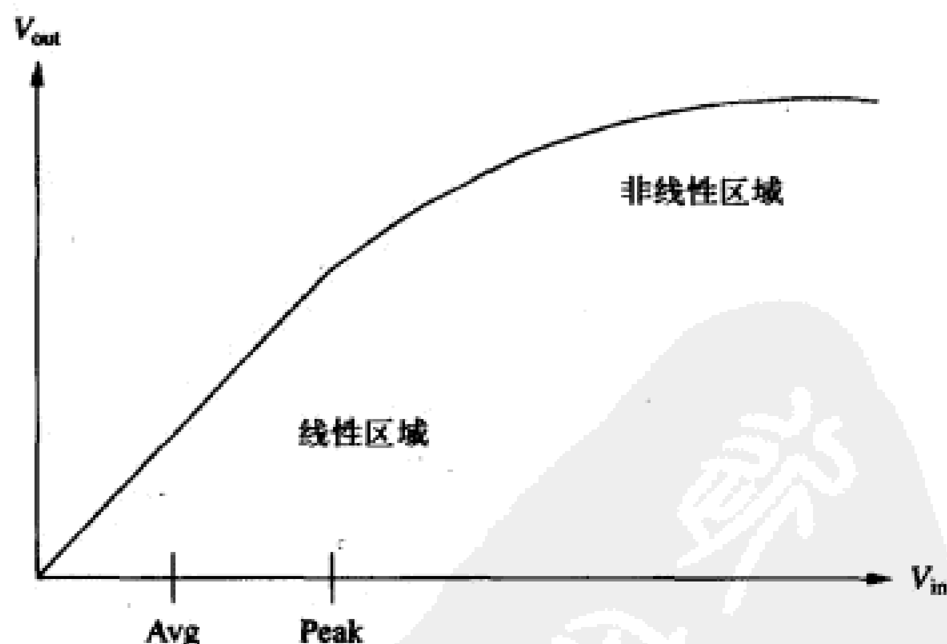


图12-11　典型功放的特性

连续时间信号的峰均比为

$$\mathrm{PAR} \triangleq \frac{\max_t |x(t)|^2}{E_t[|x(t)|^2]} \tag{12-38}$$

离散时间信号的峰均比为

$$\text{PAR} \triangleq \frac{\max_n |x[n]|^2}{E_n[|x[n]|^2]} \tag{12-39}$$

幅度为常数的信号（如方波）峰均比为0dB。正弦波的峰均比是3dB，因为，$\max[\sin^2(2\pi t/T)]=1$且

$$\mathrm{E}\left[\sin^2\left(\frac{2\pi t}{T}\right)\right] = \frac{1}{T}\int_0^T \sin^2\left(\frac{2\pi t}{T}\right)\mathrm{d}t = 0.5$$

故PAR=1/0.5=2。

放大器的输入是模拟信号，因此峰均比一般应该按式（12-38）对时间连续的信号测量。式（12-38）给出的峰均比对调制中的成形脉冲$g(t)$较敏感，一般得不到简单的解析式[27]。为了便于说明，我们将主要考虑容易分析的离散时间信号的峰均比。必须要注意，这样得到的结果没考虑成形脉冲$g(t)$，有可能很不精确。 393

考虑IFFT输出的时域样值：

$$x[n] = \frac{1}{\sqrt{N}}\sum_{i=0}^{N-1} X[i]\mathrm{e}^{\mathrm{j}2\pi in/N} \quad 0 \leqslant n \leqslant N-1 \tag{12-40}$$

$x[n]$的实部和虚部都是和式，若N足够大，则根据中心极限定理，$x[n]$是零均值的复高斯随机变量。对于较大的子载波数（$N \geqslant 64$），将IFFT的输出近似为高斯随机变量是比较精确的。对于复高斯的$x[n]$，OFDM符号的包络服从方差为σ_n^2的瑞利分布，相位服从均匀分布。瑞利分布有无限的拖尾，故信号的峰值将以非零的概率大于任意值。可以证明[28]，式（12-39）给出的峰均比大于门限$P_0 = \sigma_0^2/\sigma_n^2$的概率为

$$p(\text{PAR} \geqslant P_0) = 1 - (1 - \mathrm{e}^{-P_0})^N \tag{12-41}$$

下面讨论峰均比与子载波数的关系。设有N个零均值、单位功率的独立同分布的高斯随机变量x_n $(0 \leqslant n \leqslant N-1)$，平均信号功率$E_n\left[|x[n]|^2\right]$为

$$\begin{aligned} E\left[\frac{1}{\sqrt{N}}|x_0 + x_1 + \cdots + x_{N-1}|^2\right] &= \frac{1}{N}E|x_0 + x_1 + \cdots + x_{N-1}|^2 \\ &= \frac{E|x_0^2| + E|x_1^2| + \cdots + E|x_{N-1}^2|}{N} \\ &= 1 \end{aligned} \tag{12-42}$$

所有x_i同相相加时，上式取得最大值

$$\max\left[\frac{1}{\sqrt{N}}|x_0 + x_1 + \cdots + x_{N-1}|\right]^2 = \left|\frac{N}{\sqrt{N}}\right|^2 = N \tag{12-43}$$

394

所以，有N个子载波时的最大峰均比等于N。实际上，N个符号全都同相相加的几率几乎为0，观察到的峰均比一般要比N小几个分贝。但峰均比毕竟随子载波数的增长而以近似线性的方式增大。因此，虽然为了减小循环前缀引起的开销，我们希望N越大越好，

但代价是很高的峰均比。

有许多方法可以降低OFDM信号的峰均比，或者可以容忍峰均比。这些方法包括对高于某个门限值的OFDM信号进行削峰、用互补信号抵消峰值、采用有失真的非线性功放进行纠正、或者是采用特殊的编码技术[29]。文献[30]很好地总结了其中的一些技术。

12.5.2 频率偏移和时间偏移

OFDM调制通过正交的子信道传输数据符号，正交性是靠子载波间隔$\Delta f=1/T_N$保证的。子信道在频域可能是重叠的，例如图12-12所示出的时域为矩形脉冲（频域为sinc函数）的OFDM。在实际当中，子载波的频率间隔是非理想的，Δf 并不是精确等于$1/T_N$。其原因主要有振荡器不匹配、多普勒频移以及定时同步误差等。例如载频振荡器精度为0.1ppm时，频率偏移为$\Delta f_\varepsilon\approx f_0\times0.1\times10^{-6}$。802.11a无线局域网的载频是$f_0$=5GHz，则$\Delta f_\varepsilon\approx$500Hz，这将破坏子信道的正交性，于是FFT输出的样值将包含邻近信道的干扰。下面我们对这种子载波间的干扰（intercarrier interference，ICI）进行分析。

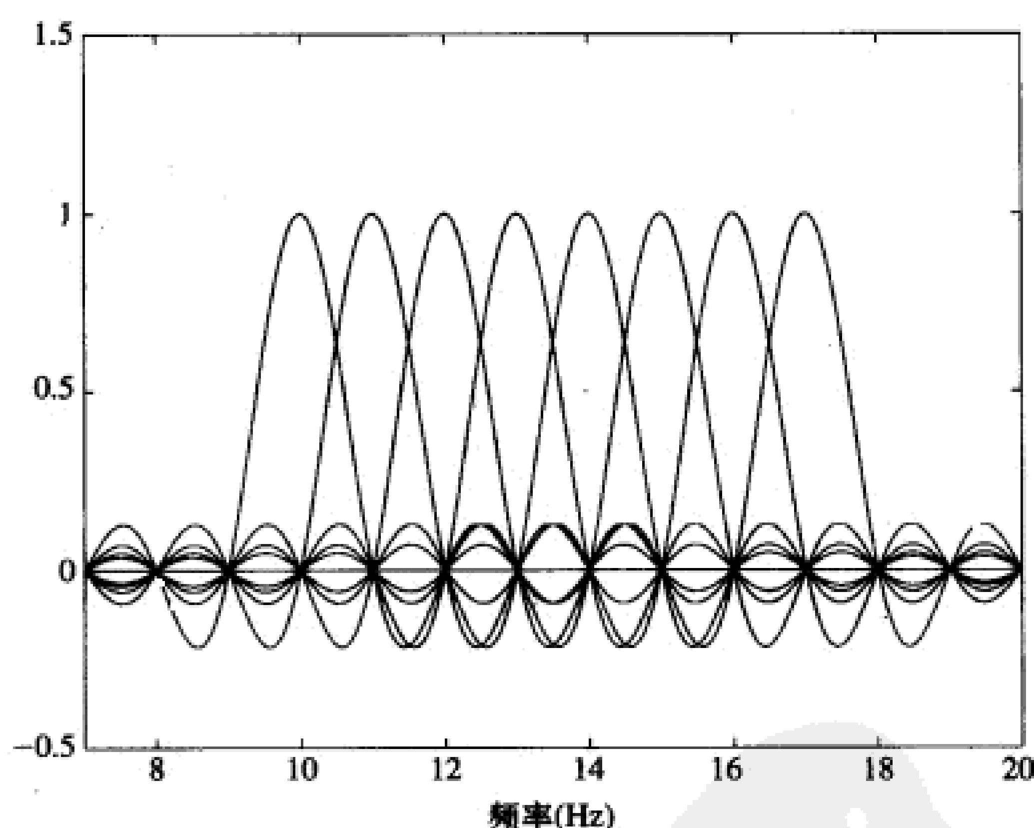


图12-12　重叠子载波的OFDM：矩形脉冲、f_0=10Hz、Δf = 1Hz

395　对于矩形脉冲成形，不考虑数据符号和载频时子载波i上的信号可以简单表示为

$$x_i(t)=\mathrm{e}^{\mathrm{j}2\pi it/T_N}\tag{12-44}$$

干扰子信道的信号可以表示为

$$x_{i+m}(t)=\mathrm{e}^{\mathrm{j}2\pi(i+m)t/T_N}\tag{12-45}$$

如果解调时的频率偏移为δ/T_N，那么这个干扰就变为

$$x_{i+m}(t)=\mathrm{e}^{\mathrm{j}2\pi(i+m+\delta)t/T_N}\tag{12-46}$$

子信道上的信号x_i和x_i+m之间的ICI是它们内积：

$$I_m = \int_0^{T_N} x_i(t)x_{i+m}^*(t)\,\mathrm{d}t = \frac{T_N(1-\mathrm{e}^{-\mathrm{j}2\pi(\delta+m)})}{\mathrm{j}2\pi(m+\delta)} \tag{12-47}$$

可以看到，$\delta=0 \Rightarrow I_m=0$。子载波$i$上总的ICI功率为

$$\mathrm{ICI}_i = \sum_{m \neq i} |I_m|^2 \approx C_0(T_N\delta)^2 \tag{12-48}$$

其中C_0是常数。从这个简单近似式中可以看出几个重要的规律。首先，当T_N增大时，子载波变窄，相互间隔也变小，于是ICI也就越大。其次，ICI随频率偏移δ按平方关系增大。另一个值得注意的地方是，式（12-48）没有直接出现N，但选择大的N值一般会使T_N也增大，从而使子载波更靠近。因此，除了峰均比随N增大这个因素外，ICI也是我们要尽量选择较小的N的原因，当然前提是要保证系统开销的要求。给定N时，采取非矩形窗能进一步降低ICI[31~32]。

定时偏移的影响一般没有频率偏移明显[5，第5章]，只要OFDM接收所用的N个样值没有受到前后符号的干扰（前提是循环前缀的长度$\mu>>\sigma_{T_m}/T_s$，σ_{T_m}是信道均方时延扩展）。可以证明，接收定时偏移τ引起的子载波i上的ICI功率近似为$2(\tau/T_N)^2$。一般$\tau<<T_N$，所以定时偏移的影响可以忽略不计。

12.6 个案研究：IEEE 802.11a无线局域网标准

IEEE 802.11a无线局域网标准就是基于OFDM技术的[6]，它占用了5GHz开放频段中的300MHz带宽。IEEE 802.11g标准在链路层设计上基本等同于IEEE 802.11a标准，只是工作在较拥挤也较窄的2.4GHz ISM开放频段上[7]。本节研究OFDM的设计及相关的设计原则。

396

802.11a有$N=64$个子载波，其中只有48个用于数据传输，其他子载波中12个置零以降低邻信道干扰，其余4个发送导频以进行信道估计。循环前缀有$\mu=16$个样值，因此每个OFDM符号总的样值数，包括数据样值和循环前缀一共是80个。编码数据打包发送，发送端定期地收到接收端反馈回来的误包率，并据此决定出合适的纠错码以及调制方式。所有的子载波在同一时间内都必须采用相同的编码及调制方式。纠错码为卷积码，有三种可能的编码率：$r=1/2$、2/3或3/4。子信道可采用的调制方式为BPSK、QPSK、16QAM或64QAM。

300MHz的总带宽分为许多20MHz的信道，信道分配给不同的用户。20MHz的信道带宽B（即取样率$1/T_s$）平分为64个子载波，所以每个子载波的带宽为：

$$B_N = \frac{20\ \mathrm{MHz}}{64} = 312.5\ \mathrm{kHz}$$

由$\mu=16$及$1/T_s=20\mathrm{MHz}$可得到能够消除ISI的最大时延扩展为

$$T_m < \mu T_s = \frac{16}{20\ \mathrm{MHz}} = 0.8\ \mu\mathrm{s}$$

这个值对应室内环境的时延扩展。包括循环前缀在内的每个OFDM符号周期有

80=64+16个采样点，因此子信道上的码元周期为

$$T_N = 80T_s = \frac{80}{20\times10^6} = 4\mu s$$

每个子信道的数据速率为$\log_2 M/T_N$。只有48个子载波用于数据，因此采用BPSK调制（1比特/符号）、r=1/2编码时的最小数据速率为

$$R_{\min} = 48\text{ subcarriers}\cdot\frac{1/2\text{ bit}}{1\text{ coded bit}}\cdot\frac{1\text{ coded bit}}{\text{subcarrier symbol}}\cdot\frac{1\text{ subcarrier symbol}}{4\cdot10^{-6}\text{ seconds}} \quad (12\text{-}49)$$
$$= 6\text{ Mbit/s}$$

最大的数据速率为

$$R_{\max} = 48\text{ subcarriers}\cdot\frac{3/4\text{ bit}}{1\text{ coded bit}}\cdot\frac{6\text{ coded bits}}{\text{subcarrier symbol}}\cdot\frac{1\text{ subcarrier symbol}}{4\cdot10^{-6}\text{ seconds}} \quad (12\text{-}50)$$
$$= 54\text{ Mbit/s}$$

397 可能的数据速率介于最大最小速率之间这个很宽的范围内。

例12.6：试求802.11a系统采用16QAM调制及2/3率编码时的数据速率。

解：16QAM调制时，每个子载波每符号传输$\log_2(16)$=4个编码比特。总共有48个子载波用于传输数据。对于2/3率的编码，每个编码比特包含2/3个信息比特。因此数据速率为

$$R_{\max} = 48\text{subcarriers}\cdot\frac{2/3\text{bit}}{1\text{codedbit}}\cdot\frac{4\text{codedbits}}{\text{subcarriersymbol}}\cdot\frac{1\text{subcarrier symbol}}{4\cdot10^{-6}\text{seconds}} \quad (12\text{-}51)$$
$$= 32\text{Mbit/s}$$

习题

12.1 证明：为了使子载波$\{\cos(2\pi j\Delta f t+\phi), j=1,2,\cdots\}$在$[0, T_N]$内形成一组正交基函数，载波间隔$\Delta f$在任意初相$\phi$的情况下至少是$1/T_N$。在初相都为零的情况下，子载波间隔可以减少一半。

12.2 设一个OFDM系统工作在相干带宽B_c=10kHz的信道上，

(a) 假设子信道上的码元周期为$T_N=1/B_N=10T_m$可以保证子信道经历平衰落，其中$T_m=1/B_c$，求子信道的码元周期。

(b) 假设系统的载波数是N=128，采用滚降系数为β=1.5升余弦脉冲。由于时间受限的原因，需要额外增加一些带宽来保证带外功率非常小，假设这个额外带宽因子为ϵ=0.1，那么系统的总带宽是多少？

(c) 求载波间隔为$1/T_N$的重叠子信道系统所需的总带宽，并与（b）中的结果进行比较。

12.3 由DFT的定义证明：离散时间序列的循环卷积对应其DFT变换的乘积。

12.4 一数据信号的带宽为0.5MHz，数据速率为0.5Mbit/s。信号通过时延扩展为10μs的无线信道传输。

(a) 如果采用非重叠子信道的多载波调制来消除ISI，大约需要多少个子载波？每个子载波的数据速率和码元周期是多少？（不必完全消除ISI，因此可以认为$T_s=T_m$就足够了。）

下面两小题假设每个子载波经历瑞利平衰落，因此ISI被完全消除。并假设第n个子信道

上的平均接收信噪比（γ_i）等于$1000/n$（线性单位）。

（b）设每个子载波都采用BPSK调制，且采用跨子载波的重复码（每个比特同时在所有子载波上发送），那么大数判决译码的误码率是多少？系统的数据速率是多少？

398

（c）设采用某种自适应加载技术（每个子载波可以用不同的星座），使得每个子载波的平均误比特率不超过10^{-3}（这里指对衰落平均后的性能，不要假设收发可以根据瞬时的衰落值自适应调整发送功率或传输速率）。求能满足这个平均误比特率要求时各子载波上的MQAM星座，并求整个系统的传输速率。

12.5　一多载波调制系统有三个不重叠的子信道，载频间隔是200kHz，每个子载波的基带带宽为100kHz。

（a）信道相干带宽为多少时，此多载波系统的子信道为平衰落（近似无ISI）？信道相干带宽为多少时，子信道上的衰落是相互独立的？如果子信道上的衰落相关，对跨信道的信道编码有什么影响？

（b）设总的发送功率为P=300mW，每个子信道的噪声功率为1mW。各子信道都以100mW发送，相应的子信道信噪比为γ_1=11dB、γ_2=14dB、γ_3=18dB。设子信道没有衰落，因而这些信噪比数值是恒定的。若要求子信道的误比特率为10^{-3}，请给出每个子信道可传输的最大MQAM星座。假设星座点数限制为2的整幂，并用上界$\mathrm{BER}\leqslant 0.2e^{-1.5\gamma/(M-1)}$来计算误比特率。假设每个子信道的码元速率为$T_s=1/3$，$B$是子信道的基带带宽，请给出该多载波信号的数据速率。

（c）对于（b）小题中的子信道的信噪比，假设采用预编码使每个子信道的接收信噪比相等，然后用相同的信号星座图发送。如欲达到（b）小题中的数据速率，应采用多大的星座图？采用这种星座图时，为使各子信道的接收信噪比能满足目标误比特率10^{-3}的要求，与（b）小题中的总发送功率300mW相比，总发送功率需要增加多少？

12.6　设信道的冲激响应为

$$h(t)=\alpha_0\delta(t)+\alpha_1\delta(t-T_1)+\alpha_2\delta(t-T_2)$$

其中T_1=10μs、T_2=20μs。为此信道设计一个多载波系统中，子信道的带宽为$B_N=B_c/2$。假设采用β=1的升余弦脉冲，并选取能使子载波正交的最小子载波间隔，那么有8个子载波的系统总占用的带宽是多少？假设每个子载波的信噪比恒为20dB，若目标误比特率要求是10^{-3}，请给出子信道上可传输的最大MQAM星座（假设星座点数必须是2的整幂），并求相应的系统总数据速率。

12.7　对于采用循环前缀的OFDM，证明式（12-22）和式（12-24）的矩阵表示是等价的。

399

12.8　证明：对$x[n]$的DFT操作可以表示为矩阵乘法$X[i]=Qx[n]$，其中

$$Q=\frac{1}{\sqrt{N}}\begin{bmatrix}1 & 1 & 1 & \cdots & 1\\ 1 & W_N & W_N^2 & \cdots & W_N^{N-1}\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 1 & W_N^{N-1} & W_N^{2(N-1)} & \cdots & W_N^{(N-1)^2}\end{bmatrix} \qquad (12\text{-}52)$$

$W_N=e^{-j2\pi/N}$。

12.9　本题表明DFT矩阵Q的行是H的特征向量。

（a）证明Q的第一行是H的特征值 $\lambda_0=\sum_{i=0}^{\mu}h_i$ 所对应的特征向量。

（b）证明Q的第二行是H的特征值 $\lambda_1=\sum_{i=0}^{\mu}h_iW_N^i$ 所对应的特征向量。

（c）用数学归纳法证明，类似的关系对Q的各行都成立。

12.10 证明：对OFDM符号添加全零后缀，再按图12-8那样把收到的拖尾加回，所得到的接收序列与采用循环前缀时相同。

12.11 考虑一个子信道个数为N=8的OFDM系统，信道的离散时间有限冲激响应为$h[n]=0.7+0.5\delta[n-1]+0.3\delta[n-3]$。

（a）写出对应式（12-23）矩阵表示$\boldsymbol{y}=\boldsymbol{Hx}+\boldsymbol{v}$的矩阵$\boldsymbol{H}$。

（b）写出对应式（12-25）矩阵表示的循环卷积矩阵$\boldsymbol{H}$，以及它的特征值分解$\boldsymbol{H}=\boldsymbol{M}\boldsymbol{\Lambda}\boldsymbol{M}^H$。

（c）对于（b）小题中的矩阵表示，每个子信道的平衰落信道增益是多少？

12.12 考虑一个5抽头的离散时间信道

$$H(z)=1+0.6z^{-1}+0.7z^{-2}+0.3z^{-3}+0.2z^{-4}$$

假设该信道模型决定了信道的最大时延扩展，并设向量编码系统的子信道数为N=256。

（a）前缀个数μ取何值时可以消除向量码字间的ISI？相应的系统开销是多少？

（b）给出该系统形如（12-23）的矩阵表示，并求信道矩阵$\boldsymbol{H}$的奇异值。

（c）求正交化子信道所需的发送预编码矩阵$\boldsymbol{V}$和接收成形矩阵$\boldsymbol{U}^H$。

12.13 分别求出β=0、0.5、1的升余弦脉冲的峰均比。β取何值时峰均比最小？这个余弦脉冲形状对定时误差是否敏感？

12.14 试求出ICI表达式（12-48）中的常数C_0。

12.15 假设利用了盲信道估计技术后，802.11a中用于导频估计的4个子信道可用于数据传输。沿用相同的调制编码，那么增加了这些子信道后，系统最大和最小的数据速率是多少？

12.16 若802.11a系统的48个可用子信道中有一半采用BPSK调制、1/2率信道编码，其余采用64QAM调制、3/4率信道编码，试求系统的数据速率。 [400]

参考文献

[1] J. Bingham, "Multicarrier modulation for data transmission: An idea whose time has come," *IEEE Commun. Mag.*, pp. 5–14, May 1990.

[2] H. Sari, G. Karam, and I. Jeanclaude, "Transmission techniques for digital terrestrial TV broadcasting," *IEEE Commun. Mag.*, pp. 100–9, February 1995.

[3] J. S. Chow, J. C. Tu, and J. M. Cioffi, "A discrete multitone transceiver system for HDSL applications," *IEEE J. Sel. Areas Commun.*, pp. 895–908, August 1991.

[4] J. M. Cioffi, "A multicarrier primer," Stanford University/Amati T1E1 contribution, I1E1.4/91–157, November 1991.

[5] A. R. S. Bahai, B. R. Saltzberg, and M. Ergen, *Multi-Carrier Digital Communications – Theory and Applications of OFDM*, 2nd ed., Springer-Verlag, New York, 2004.

[6] IEEE 802.11a-1999: *High-Speed Physical Layer in the 5 GHz Band*, 1999.

[7] IEEE 802.11g-2003: *Further Higher-Speed Physical Layer Extension in the 2.4 GHz Band*, 2003.

[8] IEEE 802.16a-2001: *IEEE Recommended Practice for Local and Metropolitan Area Networks*, 2001.

[9] C. Eklund, R. B. Marks, K. L. Stanwood, and S. Wang, "IEEE Standard 802.16: A technical

overview of the WirelessMAN 326 air interface for broadband wireless access," *IEEE Commun. Mag.*, pp. 98–107, June 2002.

[10] M. Corson, R. Laroia, A. O'Neill, V. Park, and G. Tsirtsis, "A new paradigm for IP-based cellular networks," *IT Professional*, pp. 20–9, November/December 2001.

[11] W. Lu, "4G mobile research in Asia," *IEEE Commun. Mag.*, pp. 104–6, March 2003.

[12] T. S. Rappaport, A. Annamalai, R. M. Buehrer, and W. H. Tranter, "Wireless communications: Past events and a future perspective," *IEEE Commun. Mag.*, pp. 148–61, May 2002.

[13] J. M. Cioffi. *Digital Communications*, chap. 4: *Multichannel Modulation*, unpublished course notes, available at ⟨http://www.stanford.edu/class/ee379c/⟩.

[14] S. Kasturia, J. Aslanis, and J. Cioffi, "Vector coding for partial response channels," *IEEE Trans. Inform. Theory*, pp. 741–62, July 1990.

[15] L. J. Cimini, "Analysis and simulation of a digital mobile channel using orthogonal frequency division multiplexing," *IEEE Trans. Inform. Theory*, pp. 665–75, July 1985.

[16] P. S. Chow, J. M. Cioffi, and John A. C. Bingham, "A practical discrete multitone transceiver loading algorithm for data transmission over spectrally shaped channels," *IEEE Trans. Commun.*, pp. 773–5, February–April 1995.

[17] Z. Wang, X. Ma, and G. B. Giannakis, "OFDM or single-carrier block transmissions?" *IEEE Trans. Commun.*, pp. 380–94, March 2004.

[18] R. K. Jurgen, "Broadcasting with digital audio," *IEEE Spectrum*, pp. 52–9, March 1996.

[19] S. Kaider, "Performance of multi-carrier CDM and COFDM in fading channels," *Proc. IEEE Globecom Conf.*, pp. 847–51, December 1999.

[20] P. Hoeher, S. Kaiser, and P. Robertson, "Two-dimensional pilot-symbol-aided channel estimation by Wiener filtering," *Proc. IEEE Internat. Conf. Acous., Speech, Signal Proc.*, pp. 1845–8, April 1997.

[21] A. Scaglione, G. B. Giannakis, and S. Barbarossa, "Redundant filterbank precoders and equalizers. I: Unification and optimal designs," *IEEE Trans. Signal Proc.*, pp. 1988–2006, July 1999.

[22] A. Scaglione, G. B. Giannakis, and S. Barbarossa, "Redundant filterbank precoders and equalizers. II: Blind channel estimation, synchronization, and direct equalization," *IEEE Trans. Signal Proc.*, pp. 2007–22, July 1999.

[23] A. V. Oppenheim, R. W. Schafer, and J. R. Buck, *Discrete-Time Signal Processing*, 2nd ed., Prentice-Hall, Englewood Cliffs, NJ, 1999.

[24] H. Sampath, S. Talwar, J. Tellado, V. Erceg, and A. Paulraj, "A fourth-generation MIMO-OFDM broadband wireless system: Design, performance, and field trial results," *IEEE Commun. Mag.*, pp. 143–9, September 2002. 401

[25] L. H. Brandenburg and A. D. Wyner, "Capacity of the Gaussian channel with memory: The multivariate case," *Bell System Tech. J.*, pp. 745–78, May/June 1974.

[26] G. L. Stuber, J. R. Barry, S. W. McLaughlin, Y. Li, M. A. Ingram, and T. G. Pratt, "Broadband MIMO-OFDM wireless communications," *Proc. IEEE*, pp. 271–94, February 2004.

[27] H. Ochiai and H. Imai, "On the distribution of the peak-to-average power ratio in OFDM signals," *IEEE Trans. Commun.*, pp. 282–9, February 2001.

[28] D. J. G. Mestdagh and P. M. P. Spruyt, "A method to reduce the probability of clipping in DMT-based transceivers," *IEEE Trans. Commun.*, pp. 1234–8, October 1996.

[29] K. G. Paterson and V. Tarokh, "On the existence and construction of good codes with low peak-to-average power ratios," *IEEE Trans. Inform. Theory*, pp. 1974–87, September 2000.

[30] J. Tellado, *Multicarrier Modulation with Low PAR: Applications to DSL and Wireless*, Kluwer, Boston, 2000.

[31] C. Muschallik, "Improving an OFDM reception using an adaptive Nyquist windowing," *IEEE Trans. Consumer Elec.*, pp. 259–69, August 1996.

[32] A. Redfern, "Receiver window design for multicarrier communication systems," *IEEE J. Sel. Areas Commun.*, pp. 1029–36, June 2002. 402

第13章 扩　　频

虽然带宽在无线通信中非常宝贵，但在有些时候，通过增加发射信号带宽可以提高系统的性能。和数据通信中所需的最小带宽相比，扩频技术大大扩展了信号的带宽。需要进行扩频的原因有很多。扩频可以使信号隐蔽在背景噪声下，使其难以被发现。扩频可以消除码间干扰和窄带干扰造成的性能恶化。采用RAKE接收机接收的扩频系统可以把不同的多径信号进行相干合并。相互叠加的扩频信号在解调时干扰很小，所以扩频可以使多个用户共享相同的频带。此外，宽带的扩频信号对定位和获取定时信息非常有用。

扩频通信最初应用在军事中，主要因为它能将信号隐蔽在背景噪声下，以及它能抑制窄带干扰和它不易被被对方发现或截获的特性。在商用方面，扩频的抗窄带干扰特性使其常用于无绳电话中。它抵抗ISI及多用户共享频带的特性使其适用于蜂窝系统和无线局域网中。最终，它已成为第二代、第三代蜂窝系统及第二代无线局域网的技术基础。

13.1 扩频原理

扩频是一种数字调制技术，其发送信号的带宽远大于所传信息需要的带宽。有许多技术可以使信号带宽大于最小所需，例如编码及调频，但这些技术不属于扩频。扩频调制具有如下三个特性：

- 信号带宽远大于所传信息需要的带宽。
- 扩频调制用*扩频码*（spreding code），扩频码独立于所传数据。
- 接收端的解扩是用同步扩频码和接收信号进行相关。

403

为了更清楚地说明这一点，我们用5.1节中介绍的信号空间概念来看扩频，它把带宽为B的信息信号加载到一个带宽为B_s的发射信号上，B_s远大于需要的带宽B。根据5.3节，一组带宽为B、持续时间为T的线性无关信号$s_i(t)$，$i=1, 2, \cdots, M$ 可以表示为：

$$s_i(t)=\sum_{j=1}^{N} s_{ij}\phi_j(t) \qquad 0\leqslant t<T \tag{13-1}$$

其中的正则基函数$\phi_j(t)$张成了一个N维空间。每T秒发射一个信号，每秒的信息量为$\log_2 M/T$比特。根据5.1.2节，表示这些信号大约需要$2BT$个基函数。由于$\{s_i(t)\}_{i=1}^{M}$线性无关，所以$M\approx 2BT$。我们选择$N>>M$以将这些信号放入到高维空间中。接收端有M个支路，第i个支路用$s_i(t)$对接收信号进行相关运算。接收机输出各支路相关最大者。

假设用随机序列产生信号$s_i(t)$，则系数s_{ij}由随机序列决定，其均值为0、方差为E_s/N。因此信号$s_i(t)$的能量均匀分布在信号空间的N个维上。考虑信号空间中的一个干扰，它可以表示为

$$I(t)=\sum_{j=1}^{N} I_j\phi_j(t) \tag{13-2}$$

$[0, T]$内的总能量为

$$\int_0^T I^2(t)\,\mathrm{d}t = \sum_{j=1}^{N} I_j^2 = E_J \tag{13-3}$$

假设发送$s_i(t)$，忽略噪声时的接收信号是发送信号和干扰之和：

$$x(t) = s_i(t) + I(t) \tag{13-4}$$

接收端第i个相关器的输出为

$$x_i = \int_0^T x(t)s_i(t)\,\mathrm{d}t = \sum_{j=1}^{N}(s_{ij}^2 + I_j s_{ij}) \tag{13-5}$$

式中第一项为有用信号，第二项为干扰。可以证明[1]，信干比（signal-to-interference power ratio，SIR）为：

$$\mathrm{SIR} = \frac{E_s}{E_j}\cdot\frac{N}{M} \tag{13-6}$$

404

这个结果与干扰在N维空间中的能量分布无关。它表明，将干扰的功率扩展到比信号维度M更大的N维空间后，信干比增加了$G=N/M$倍，称G为*处理增益*（processing gain）或者*扩频因子*（spreading factor）。实际扩频系统的处理增益在100~1000量级。由于$N\approx 2B_sT$，$M\approx 2BT$，所以$G\approx B_s/B$，即扩频信号带宽与信息信号带宽的比值。处理增益一般定义为带宽比或者类似比值，但其内在含义是指干扰条件下扩频系统相对于未扩频系统的性能增益（见文献[2]的2.1节）。注意在噪声和干扰环境下，分组码和卷积码也通过增加信号带宽而提高了系统性能。给定扩展带宽时，到底用编码好还是用扩频好是一个需要权衡的问题，取决于具体系统的设计要求[3]。

扩频的实现一般有两种形式：*直接序列*（direct sequence, DS）及*跳频*（frequency hopping，FH）。直序扩频（direct-sequence spread-spectrum，DSSS）调制将已调的数据信号$s(t)$与一个*扩频信号*（spreading signal）或*扩频码*（spreading code）$s_c(t)$相乘，其中$s_c(t)$在时间T_c内的取值是固定的1或者－1。一般称扩频码的比特为*码片*（chip），称T_c为*码片时间*（chip time），称$1/T_c$为*码片速率*（chip rate）。$s_c(t)$的带宽$B_c\approx 1/T_c$近似是已调信号$s(t)$带宽B的$B_c/B\approx T_s/T_c$倍，每比特中的码片数是一个约等于G的整数。已调信号和扩频信号相乘的结果是频域卷积。因此，发送信号$s(t)\,s_c(t)$的频域响应为$S(f)*S_c(f)$，其带宽约为$B+B_c$。图13-1示出了一个码元周期内，扩频信号$s_c(t)$与已调信号$s(t)$的相乘过程。

在AWGN信道中，接收的扩频信号为$s(t)s_c(t)+n(t)$，$n(t)$是信道噪声。将此接收信号乘以同步的扩频信号$s_c(t)$，得到$s(t)s_c^2(t)+s_c(t)n(t)$。由$s_c(t)=\pm 1$得$s_c^2(t)=1$。此外，如果$s_c(t)$均值为0，且带宽足够宽（即自相关函数近似为冲激函数），那么$n'(t)=s_c(t)n(t)$与噪声$n(t)$有相同的统计特性。这样，接收信号是$s(t)s_c^2(t)+s_c(t)n(t)=s(t)+n'(t)$，说明扩频和解扩对信号在AWGN信道中的传输没有产任何影响。然而，当信道中有窄带干扰和ISI时，扩频和解扩有很大的好处。

现在我们先从频域上说明直序扩频抑制窄带干扰和多径干扰的特性，后面再进行详细讨论。对窄带干扰的抑制如图13-2所示。不考虑噪声，接收机的输入由扩频信号$S(f)*S_c(f)$和窄带干扰$I(f)$构成。接收端经过解扩后得到信号$S(f)$。而干扰信号$I(t)$与扩频信

号$s_c(t)$相乘后在频域是卷积$S_c(f)*I(f)$，因此解扩的效果就是把干扰信号的功率分布到扩频信号的带宽上。对$s(t)$的解调等效于一个低通滤波器，它滤除了扩频后的大部分干扰功率，输出的干扰功率降低了$G\approx B_c/B$倍。

405

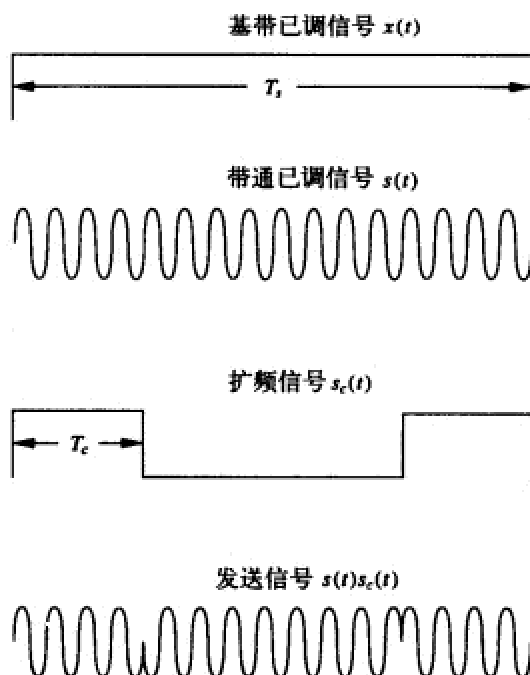


图13-1 扩频信号相乘

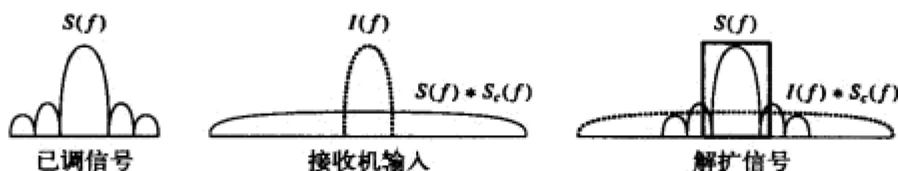


图13-2 直序扩频抗窄带干扰

对ISI的抑制与此类似，如图13-3所示。假设扩频信号为$s(t)s_c(t)$通过一个冲激响应为$h(t)=\alpha\delta(t)+\beta\delta(t-\tau)$的两径信道传输。信道的频域响应为$H(f)=\alpha+\beta e^{-j2\pi f\tau}$，忽略噪声时的接收机输入在频域为$H(f)[S(f)*S_c(f)]$，在时域为$[s(t)s_c(t)]*h(t)=\alpha s(t)s_c(t)+\beta s(t-\tau)s_c(t-\tau)$。假设接收端解扩时所乘的信号是对两径模型中的第一径同步的$s_c(t)$，则时域结果是$\alpha s(t)\,s_c^2(t)+\beta s(t-\tau)s_c(t-\tau)s_c(t)$。由于第二径分量$\beta s'(t)=\beta s(t-\tau)s_c(t-\tau)s_c(t)$是与异步的$s_c(t)$相乘，它还是一个扩频信号，解调器能滤除其大部分能量。从13.2节可知，解调过程将对多径成分引入一个衰减因子$\rho(\tau)$，它是扩频码对应延迟τ的自相关。当$\tau>T_c$时，这个自相关非常小，在$1/G\approx T_c/T_s$量级，从而使扩频信号能显著消除ISI。由于扩频码的自相关特性决定了扩频系统抑制ISI的能力，所以采用有良好的自相关性的扩频码就非常重要。下一节将具体讨论扩频码的设计。

406

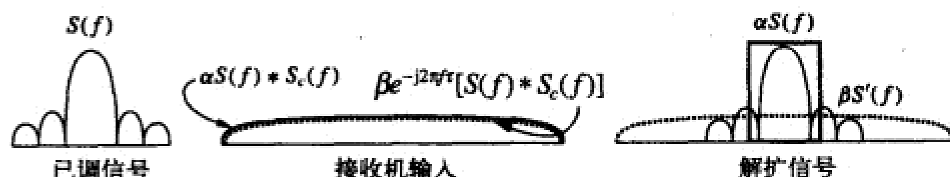


图 13-3

跳频扩频[1]（frequence-hopping spread spectrum，FHSS）的基本思想是在一个宽广的频带内，根据扩频码$s_c(t)$的值不断改变已调信号的载频，如图13-4所示。$s_c(t)$的码片周期T_c是两次跳变时间间隔，在此时间内已调数据信号的载频f_i不变。跳变时间可以大于码元周期，即$T_c=kT_s$，k为整数，此时称为为慢跳频（slow frequency hopping，SFH）。也可以在一个码元周期内跳多次，即$T_c=T_s/k$，此时称为为快跳频（fast frequency hopping，FFH）。快跳频的每个符号中都有频率分集，它能抑制窄带干扰，并能抵抗频率选择性衰落引起的频谱零点。跳频系统的带宽近似为NB，其中N为可跳的载频数，B是数据信号的带宽。跳频系统一般采用某种FM调制（如CPFSK），所需载频用一个受码片序列控制的频率合成器产生。接收端用相同的频率合成器解调，这个合成器所用的码片序列与$s_c(t)$同步。如同直序扩频，跳频能对抗窄带干扰和多径，但在AWGN信道中不起作用。

407

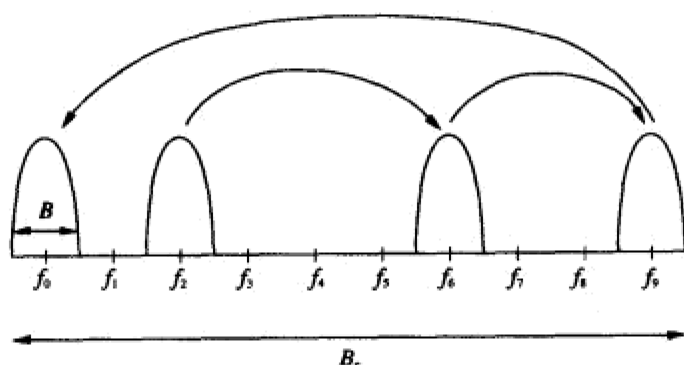


图 13-4

设有一个窄带干扰，其带宽为B，频率在跳频系统的一个频率f_i上。只有当跳频序列产生的频率是f_i时，FH信号才和干扰在相同频带上。如果跳频序列使各个频率的使用时间相同，则干扰仅在$1/N$的时间内存在，因此，干扰功率被减小到了原来的$1/N$。不过DS和FH降低干扰的机理不同。DS在所有时间上都降低了干扰功率，而FH在一部分时间内干扰还是满功率的。FFH系统中干扰只在码元的一部分时间内存在，所以不用编码就能对抗干扰。在SFH系统中，干扰会影响多个码元，所以一般要采用编码及交织技

1. 跳频的概念是二战期间由电影明星Hedy Lamarr和作曲家George Antheil发明的。他们的专利"秘密通信系统"用自动钢琴卷帘谱产生的码片序列在88个音调上跳变。设计目的是为了无线电制导的鱼雷不易被发现或被干扰。

术以避免一个码字中同时有多个错误。FH常用在军事通信中，那里的干扰源往往是旨在破坏通信系统的恶意干扰。

再来看多径对FH系统的影响。简单起见，我们考虑多径时延为τ的两径信道。假设接收机同步于直射径的跳频序列。在相应的频率上，直射径被解调，而多径分量在时延τ之后才到达。若$\tau>T_c$，那么当频率为f_i的第二径信号到达时，接收机已经跳到一个新的载频$f_j\neq f_i$上，在那里进行下变频和解调，由于频率不同，所以第二条径对解调信号的影响可以忽略。因此，$\tau>T_c$时，解调信号既无平衰落也无频率选择性衰落。$\tau<T_c$时，多径的影响与已调数据信号的带宽B和跳变率有关。对于$T_c<<T_s$的FFH系统，因为已假设$\tau<T_c$，故有$\tau<T_c<<T_s$，所有多径分量都在一个码元间隔内到达，多径的效果是一个复数增益，信号经历平衰落。对于$T_c>>T_s$的SFH系统，已假设$\tau>T_s$，故所有多径分量到达接收机时，信号频率还没有变。此时若$B<1/\tau$，信号经历平衰落，若$B>1/\tau$，信号经历频率选择性衰落。等效基带信道与频率有关，载频跳变时，等效基带信道也要发生变化，因此所经历的衰落信道也在慢变化。综上所述，$\tau>T_c$时，跳频消除了多径对直射径解调的影响。$\tau<T_c$时，FFH系统呈现出平衰落，而SFH系统当$B<1/\tau$时呈现为慢变平衰落，当$B>1/\tau$时呈现为慢变频率选择性衰落。时变平衰落或频率选择性衰落下的性能分析与6.3节、6.5节所讨论的无跳频系统相同。

408

除了可以抵抗干扰和ISI外，DSSS和FHSS都可以实现多址接入，允许多个用户以很小的干扰共享扩频带宽。在这样的多用户系统中，用户间的干扰取决于它们扩频码的互相关特性。扩频码的设计一般是要么具有良好的自相关特性以减小ISI，要么具有良好的互相关特性能以减小多用户干扰。经常需要在这两个特性之间权衡，最佳的选择与用户数、多径及干扰的严重程度等有关。跳频和直接序列技术的权衡将在13.5节中讨论。跳频也用在蜂窝系统中以平均邻小区干扰。

例13.1：某SFH系统的跳频时间为$T_c=10\mu s$，码元间隔为$T_s=1\mu s$。跳频信号通过多径信道传输。请问，当多径时延扩展大约在什么范围时，接收端解扩的信号呈现出频率选择性衰落？

解：根据前面对两径信道的分析，仅当时延扩展$\tau<T_c=10\mu s$时，信号呈现出衰落（平衰落或频率选择性衰落）。此外，频率选择性衰落要求$B\approx 1/T_s=10^6>1/\tau$，即要求$\tau>10^{-6}=1\mu s$。所以当时延扩展大约在$1\mu s$~$10\mu s$范围内时，解扩信号将呈现出频率选择性衰落。

13.2 直序扩频（DSSS）

13.2.1 DSSS系统模型

图13-5是端到端的DSSS系统框图。图中乘$s_c(t)$和乘载波$\cos(2\pi f_c t)$的次序可以调换。先下变频后再解扩时，码同步和解扩可以数字化实现，但使载波相位跟踪带来了困难，

因为载波跟踪要对宽带的扩频信号进行[1]。简单起见，接收端只示出了同相信号，正交信号结构与此类似。数据符号s_l先经过线性调制成为基带调制信号 $x(t)=\Sigma_l s_l g(t-lT_s)$，$g(t)$是调制器成形脉冲，$T_s$是码元间隔，$s_l$是第$l$码元间隔内发送的符号。采用线性调制是因为DSSS自身是一种相位调制，与线性调制结合时性能最好。然后将已调制的信号与码片周期为T_c的扩频码$s_c(t)$相乘，再与载波$\cos(2\pi f_c t)$相乘进行上变频。扩频信号经过信道$h(t)$传输，信道还引入了窄带干扰$I(t)$以及功率谱密度为$N_0/2$的零均值白高斯噪声$n(t)$。

409

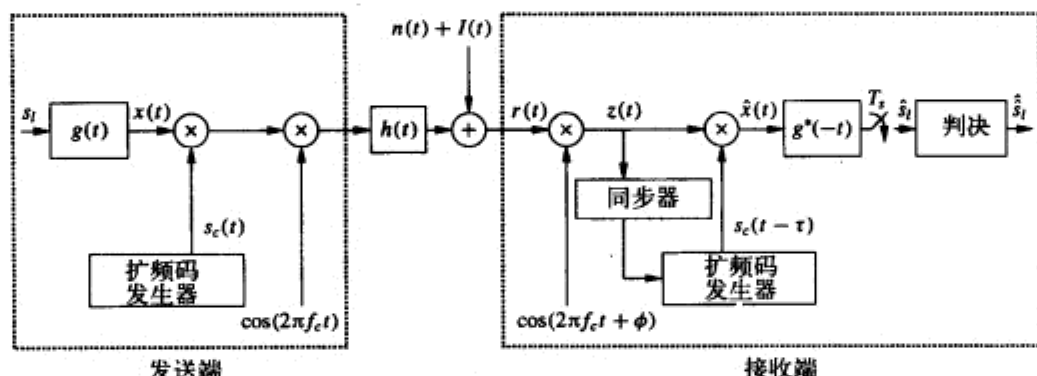


图13-5 DSSS系统模型

假设信道引入了许多多径成分：$h(t)=\alpha_0\delta(t-\tau_0)+\alpha_1\delta(t-\tau_1)+\cdots$。接收信号先经过下变频转成基带信号。假设理想载波恢复，因此接收端的载波$\cos(2\pi f_c t+\phi)$的相位ϕ与到达信号的相位匹配，即若$h(t)=\delta(t)$，则有$\phi=0$。一般假设载波恢复环路锁定在直射波或者时延最小的多径分量上。同步器用基带信号$z(t)$调整接收端扩频码发生器的时延τ，使扩频码与某一个时延为τ_i的多径分量对齐。扩频码发生器的输出是扩频码$s_c(t-\tau)$，如果同步器能理想对齐第i径的时延，则$\tau=\tau_i$。理想同步器将锁定在幅度最大的多径分量上。不过达到这一点需要复杂的搜索过程，因此实际的同步器一般锁定在最先找到的幅度高于给定门限的多径分量上。同步过程可能相当复杂，尤其在信道存在着严重ISI的时候。同步电路占扩频接收机的一大部分。13.2.3节将对同步问题做更详细的讨论。

与扩频码$s_c(t-\tau)$相乘后，时延为τ的多径分量被解扩，其他路径的信号没有被解扩，稍后我们将看到，这些分量的大部分能量也被消除。解扩后的基带信号$\hat{x}(t)$经过匹配滤波器然后判决。因此，DSSS系统的解调器由三部分组成：下变频、解扩、基带解调。这

1. 对带通已调信号进行扩频和解扩的工作过程是：发送端用标准的窄带调制器产生已调带通信号，再进行扩频。接收端先解扩，再用标准的窄带解调器解调。这样的次序可以使扩频系统直接使用现有的窄带调制解调器，并且不影响载波恢复等操作。不过当今的扩频系统尽量把所有的信号处理变成数字实现。因此扩频系统一般先在基带用数字信号处理进行码元调制并乘以扩频码，再经过A/D转换后以模拟方式上变频到载波。在这种情况下，图13-5中乘载波之前的所有部分都是数字实现的，乘$s_c(t)$之后是D/A变换。这种系统中的载波恢复设计非常困难，因为它要对扩频信号操作。尤其是，载波提取经常要用到一些非线性操作（如平方），用来去除信息或扩频序列，扩频信号中的噪声将严重恶化这些操作的性能。

410 样的，解调器也称为DSSS的单用户匹配滤波检测器。现在我们再进一步讨论这三部分。

出于简便，假设采用矩形脉冲 $g(t)=\sqrt{2/T_s}, 0 \leqslant t \leqslant T$。匹配滤波器$g^*(-t)$简化为对$\hat{x}(t)$乘以$\sqrt{2/T_s}$后在0到$T_s$内积分，积分结果就是发送符号的估计值。已假设了理想载波恢复，所以接收机的载波相位ϕ匹配于到达信号的相位。我们假设接收端的同步也是理想的。对干扰和多径的抑制出现在解调过程中。具体来说，匹配滤波器的输入为

$$\begin{aligned}\hat{x}(t) = ([x(t)s_c(t)\cos(2\pi f_c t)] * h(t))s_c(t-\tau)\cos(2\pi f_c t+\phi) \\ + n(t)s_c(t-\tau)\cos(2\pi f_c t+\phi) + I(t)s_c(t-\tau)\cos(2\pi f_c t+\phi)\end{aligned} \tag{13-7}$$

如果没有多径，则$h(t)=\delta(t)$，$\phi=0$，接收端也将理想同步于$\tau=0$。此时，扩频解扩对基带信号$x(t)$没有影响。实际上，扩频码的幅度为±1，接收端对$s_c(t)$乘以其同步副本的结果是$s_c^2(t)=1$。如果没有干扰和多径，即若$h(t)=\delta(t)$、$I(t)=0$时，则有

$$\begin{aligned}\hat{x}(t) &= x(t)s_c^2(t)\cos^2(2\pi f_c t) + n(t)s_c(t)\cos(2\pi f_c t) \\ &= x(t)\cos^2(2\pi f_c t) + n(t)s_c(t)\cos(2\pi f_c t)\end{aligned} \tag{13-8}$$

如果$s_c(t)$的带宽足够宽，可近似认为$n(t)\ s_c(t)$的统计特性与$n(t)$相同，是均值为零、功率谱密度为$N_0/2$高斯白噪声。$g(t)=\sqrt{2/T_s}, 0 \leqslant t \leqslant T_s$的匹配滤波器输出为

$$\begin{aligned}\hat{s}_l &= \int_0^{T_s} \hat{x}(t) * g^*(-t)\,dt \\ &= \sqrt{\frac{2}{T_s}}\int_0^{T_s} x(t)\cos^2(2\pi f_c t)\,dt + \sqrt{\frac{2}{T_s}}\int_0^{T_s} n(t)s_c(t)\cos(2\pi f_c t)\,dt \\ &= \frac{2}{T_s}\int_0^{T_s} s_l\cos^2(2\pi f_c t)\,dt + \sqrt{\frac{2}{T_s}}\int_0^{T_s} n(t)s_c(t)\cos(2\pi f_c t)\,dt \\ &\approx s_l + n_l\end{aligned} \tag{13-9}$$

式中假设$f_c >> 1/T_s$，s_l和n_l是无扩频的系统解调输出的数据和噪声。

现在考虑有一个载频为f_c的干扰$I(t)$，其模型为$I(t)=I'(t)\cos(2\pi f_c t)$，$I'(t)$是某个窄带基带信号。仍假设$h(t)=\delta(t)$。输入信号乘以理想同步的扩频信号后成为

$$\hat{x}(t) = x(t)\cos^2(2\pi f_c t) + n(t)s_c(t)\cos(2\pi f_c t) + I'(t)s_c(t)\cos^2(2\pi f_c t) \tag{13-10}$$

式中$n(t)\ s_c(t)$仍认为是均值为零的高斯白噪声。于是，解调输出为

411

$$\begin{aligned}\hat{s}_l &= \frac{2}{T_s}\int_0^{T_s} s_l s_c^2(t)\cos^2(2\pi f_c t)\,dt + \sqrt{\frac{2}{T_s}}\int_0^{T_s} n(t)s_c(t)\cos(2\pi f_c t)\,dt \\ &\quad + \sqrt{\frac{2}{T_s}}\int_0^{T_s} I'(t)s_c(t)\cos^2(2\pi f_c t)\,dt \\ &\approx s_l + n_l + I_l\end{aligned} \tag{13-11}$$

式中假设$f_c >> 1/T_s$，s_l和n_l是无扩频的系统解调输出的数据和噪声。从式（13-11）的最后一项可以看出对窄带干扰的抑制。$I'(t)s_c(t)$是宽带信号，带宽大约为$1/T_c$，积分相当

于一个低通滤波器，带宽大约为$1/T_s=1/T_c$，从而消除了大部分的干扰功率。

再讨论对ISI的抑制。假设一个两径信道$h(t)=\alpha_0\,\delta(t)+\alpha_1\,\delta(t-\tau_1)$，出于简便，假定$\tau_1=kT_s$是码元周期的整数倍。假设第一径比第二径强，即$\alpha_0>\alpha_1$，接收端与第一条径同步，即图13-5中$\phi$=0、$\tau$=0。这样，无窄带干扰（$I(t)=0$）时的解扩输出为

$$\begin{aligned}\hat{x}(t) &= \alpha_0 x(t)s_c^2(t)\cos^2(2\pi f_c t)\\ &\quad+\alpha_1 x(t-\tau_1)s_c(t-\tau_1)\cos(2\pi f_c(t-\tau_1))s_c(t)\cos(2\pi f_c t)\\ &\quad+n(t)s_c(t)\cos(2\pi f_c t)\end{aligned} \tag{13-12}$$

因为$\tau_1=kT_s$，所以ISI是第$(l-k)$个发送符号引起的，即$x(t-\tau_1)=x(t-kT_s)=s_{l-k}g(t-(l-k)T_s)$。第$l$个码元周期内的解调输出为

$$\begin{aligned}\hat{s}_l &= \frac{2}{T_s}\int_0^{T_s}\alpha_0 s_l\cos^2(2\pi f_c t)\,\mathrm{d}t\\ &\quad+\frac{2}{T_s}\int_0^{T_s}\alpha_1 s_{l-k}s_c(t)s_c(t-\tau_1)\cos(2\pi f_c t)\cos(2\pi f_c(t-\tau_1))\,\mathrm{d}t\\ &\quad+\sqrt{\frac{2}{T_s}}\int_0^{T_s}n(t)s_c(t)\cos(2\pi f_c t)\,\mathrm{d}t \qquad (13\text{-}13)\\ &\approx \alpha_0 s_l+\alpha_1 s_{l-k}\cos(2\pi f_c\tau_1)\rho_c(\tau_1)+n_l \qquad (13\text{-}14)\end{aligned}$$

与前面一样，式中假设$f_c>>1/T_s$，s_l和n_l是无扩频的系统解调输出的数据和噪声。式中的第二项$\alpha_1 s_{l-k}\cos(2\pi f_c\tau_1)\rho_c(\tau_1)$来自下面这个积分：

$$\begin{aligned}&\frac{2}{T_s}\int_0^{T_s}s_c(t)s_c(t-\tau_1)\cos(2\pi f_c t)\cos(2\pi f_c(t-\tau_1))\,\mathrm{d}t\\ &\quad=\frac{1}{T_s}\int_0^{T_s}s_c(t)s_c(t-\tau_1)(\cos(2\pi f_c\tau_1)+\cos(4\pi f_c t-2\pi f_c\tau_1))\,\mathrm{d}t \qquad (13\text{-}15)\\ &\quad\approx\cos(2\pi f_c\tau_1)\frac{1}{T_s}\int_0^{T_s}s_c(t)s_c(t-\tau_1)\,\mathrm{d}t\\ &\quad=\cos(2\pi f_c\tau_1)\rho_c(t,t-\tau_1)\end{aligned}$$

其中的近似基于$f_c>>T_s^{-1}$。其中的

$$\rho_c(t,t-\tau_1)\triangleq\frac{1}{T_s}\int_0^{T_s}s_c(t)s_c(t-\tau_1)\,\mathrm{d}t \tag{13-16}$$

412

定义为扩频码在码元周期内的自相关函数[1]。更一般地，假设t是T_c的整数倍，则对于给

1. 注意，如果时延τ_1不是码元周期的整数倍时，（13-14）中的第二项将更复杂。假设$g(t)=\sqrt{2/T_s}$，$\tau_1=(k+\kappa)T_s$，$0<\kappa<1$。则当$0\leqslant t\leqslant\kappa T_s$时$x(t-\tau_1)=\sqrt{2/T_s}s_{l-k-1}$，当$\kappa T_s\leqslant t\leqslant T_s$时$x(t-\tau_1)=\sqrt{2/T_s}s_{l-k}$。这样，（13-14）的第二项成为

$$\alpha_1 s_{l-k-1}\frac{\cos(2\pi f_c\tau_1)}{T_s}\int_0^{\kappa T_s}s_c(t)s_c(t-\tau_1)\,\mathrm{d}t+\alpha_1 s_{l-k}\frac{\cos(2\pi f_c\tau_1)}{T_s}\int_{\kappa T_s}^{T_s}s_c(t)s_c(t-\tau_1)\mathrm{d}t$$

其中的每一项都是按部分码元间隔积分的扩频码自相关函数。

定的时间t和时延τ，扩频码在区间$[0, T]$上的自相关函数定义为

$$\rho_c(t, t-\tau) \triangleq \frac{1}{T}\int_0^T s_c(t)s_c(t-\tau)\,\mathrm{d}t = \frac{1}{N_T}\sum_{n=1}^{N_T} s_c(nT_c)s_c(nT_c-\tau) \tag{13-17}$$

式中$N_T=T/T_c$是T时间内的码片数，第二个等式是因为$s_c(t)$在码片周期T_c内是常数。可以证明，$\rho_c(t, t-\tau)$是对称函数，它在$\tau=0$时有最大值。此外，若$s_c(t)$的周期为T，则自相关函数只与扩频码时延差有关，即

$$\frac{1}{T}\int_0^T s_c(t-\tau_0)s_c(t-\tau_1)\,\mathrm{d}t = \rho_c(\tau_1-\tau_0) \tag{13-18}$$

此时，式（13-17）中的$\rho_c(t, t-\tau) \triangleq \rho_c(\tau)$只与$\tau$有关。

根据式（13-15），如果$T=T_s$且$\rho_c(\tau)=\delta(\tau)$，就能完全消除ISI。但在实际中，有限长度的扩频码不可能有冲激自相关。因此，人们开展了大量工作，以设计出一种在码元周期内自相关函数可以近似为冲激的扩频码。下一节将讨论抑制ISI的扩频码，其中包括最大线性码，它出色的自相关特性能最大程度地减小ISI的影响。

13.2.2　抑制ISI的扩频码：随机码、伪随机码、m序列

扩频码以确定的方式产生，扩频码发生器经常采用图13-6这种带反馈的移位寄存器，它可产生一个由二进制0、1构成序列b，称为码片序列。码片序列对脉冲宽度为T_c的方波进行幅度调制，码片为1时脉冲幅度为1，码片为0时脉冲幅度为－1。这样得到的扩频码$s_c(t)$在频域是一个sinc函数，对应矩形脉冲的傅里叶变换。n级移位寄存器的输出是周期性的，最大周期为2^n-1。扩频信号$s_c(t)$不能有直流分量，直流分量会在频谱的直流处形成一个尖峰，也会使解扩后的噪声均值不为零，因此要求码片序列中1和0的个数要近似相等。我们还希望序列中的连1或连0（称为游程）尽量短。这是因为，如果码片序列中有一个连续k个1或0的游程，那么数据信号在kT_c时间内将是一个常量，这样将使频带大约降低为$1/k$，频带降低便失去了扩频的优点。理想的情形是码片大约每个码片都改变，使频谱扩展最大化。为了使ISI的影响最小，根据（13-15）式，我们要求扩频码的自相关函数近似有$\rho_c(\tau)\approx\delta(\tau)$。

413

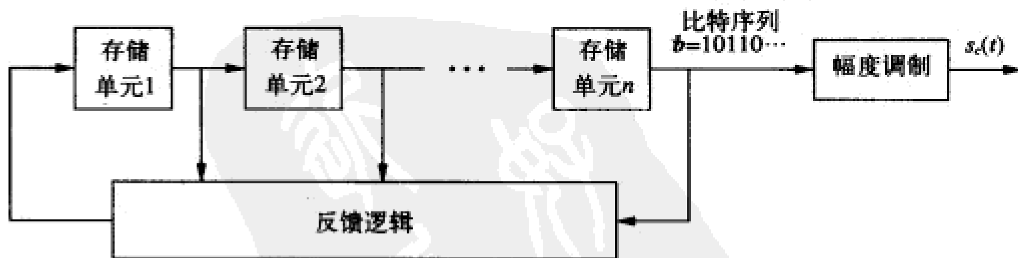


图13-6　扩频码的产生

例13.2：若码片周期为T_c=1μs，求扩频码$s_c(t)$的基带带宽。

解：扩频码$s_c(t)$由一系列的持续时间为T_c、幅度为±1的矩形脉冲组成，每个脉冲的傅里叶变换为$S(f)=T_c\,\mathrm{sinc}(fT_c)$，其主瓣的宽度是$2/T_c$。基带的零值带宽定义为使$S(f)$=0的最小值，为$1/T_c$。

DSSS中的码片序列是确定性的，为了更好地设计这个确定性的扩频序列，我们先来考察随机序列的特性。随机二进制码片序列由独立同分布的比特组成，取值0、1的概率各为0.5。长为N的随机序列比如可以通过掷币N次来产生：正面为1、反面为0。无限长随机序列的许多特性正是我们对扩频码所期望的[4]。具体来说，无限长随机序列中1和0的个数相同，我们称这样的码具有*均衡性*（balanced property）。无限长随机序列的游程一般很短，长度为1的游程个数占总游程数的一半，长度为2的游程占总游程数的四分之一，长度为r的游程占总游程数的$1/2^r$。扩频码游程长度的分布称为*游程特性*（run-length property）。随机序列还有一个特性是，经过任意移位后得到的序列和原序列相比，有一半相同，一半不同。这一点称为*移位特性*（shift property）。按照Golomb的定义[6]，如果一个确定序列在其长度趋于无限时具有均衡性、游程特性和移位特性，则称此确定序列为*伪随机序列*（pseudorandom sequency）。这三个特性对系统性能分析非常重要，因此DSSS系统经常按随机序列进行分析，这要比确定性扩频序列容易得多（见文献[5]的2.2节）。 414

在所有线性码中，m序列（即最大长度序列）具有许多良好的特性。m序列是一种循环码（见8.2.4节），因此可以用生成多项式描述并产生，其特性可用代数编码理论推出（见文献[2]的3.3节及文献[5]的2.2节）。m序列具有n级移位寄存器最大可能的周期$N=2^n-1$，序列每隔NT_c秒就重复一次。由于它是循环码，所以m序列的移位还是m序列。还有一个特性是：m序列和它的延迟进行模2加，所得到的还是一个m序列，这个m序列是原序列的延迟。这就是m序列的*移位相加特性*（shift and add property）。m序列在一个周期内1和0的个数几乎相同：0的个数为$2^{n-1}-1$，1的个数为2^{n-1}。因此由m序列生成的扩频码（也称为最大线性码）直流分量很小。m序列的游程特性与随机二进制序列几乎相同：长度为$r<n$的游程占$1/2^r$，长度为$r=n$游程占$1/2^{r-1}$。最后，用m序列的移位相加特性和均衡性可以导出，它与随机序列有相同的移位特性。m序列在均衡性、游程特性和移位特性方面和随机序列相同，因此它是一种伪随机序列（见文献[5]的2.2节）。

最大线性扩频码在周期$T=NTc$上的自相关函数为

$$\rho_c(\tau)=\begin{cases}1-|\tau|(1+1/N)/T_c & |\tau|\leqslant T_c\\ -1/N & |\tau|>T_c\end{cases} \tag{13-19}$$ 415

其中$|\tau|<(N-1)T_c$，如图13-7所示。扩频码是周期的，所以自相关函数也是周期的，且周期同为$T=NT_c$，如图13-8所示。如果τ不在码片周期kNT_c（k为任意整数）内，则$\rho_c(\tau)=-1/N=-1/(2^n-1)$。只要$n$取得足够大，那么时延$\tau$不在码片周期$kNT_c$内的多径干扰就可以基本消除。当时延$\tau$落在码片周期$kNT_c$内时，对多径干扰的衰减程度取决于自相关函数$\rho_c(\tau)$，它随着$\tau$接近$kNT_c$而线性增加。对自相关函数$\rho_c(\tau)$进行傅里叶变换可得到$s_c(t)$的功率谱密度为

$$P_{s_c}(f)=\sum_{m=-\infty}^{\infty}\frac{N+1}{N^2}\operatorname{sinc}^2\left(\frac{m}{N}\right)\delta\left(f-\frac{m}{T}\right) \tag{13-20}$$

因为$\rho_c(\tau)$是周期函数，所以$P_{s_c}(f)$是离散的，其频率分量是$1/T=1/NT_c$ Hz的倍数。

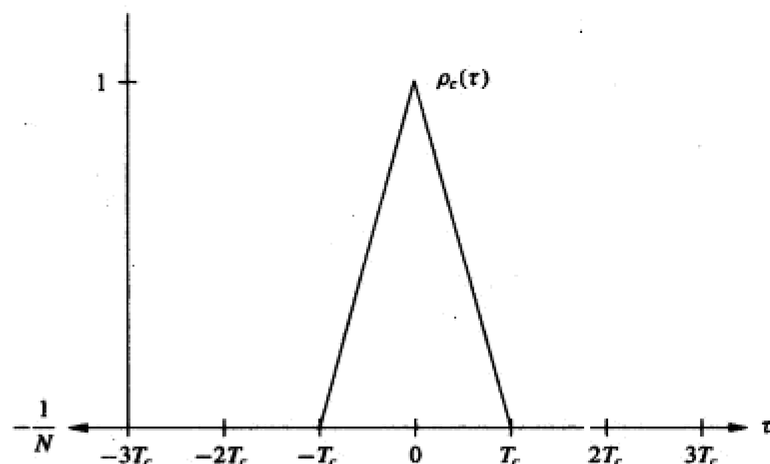


图13-7 最大线性码的自相关函数（$N=T/T_c$）

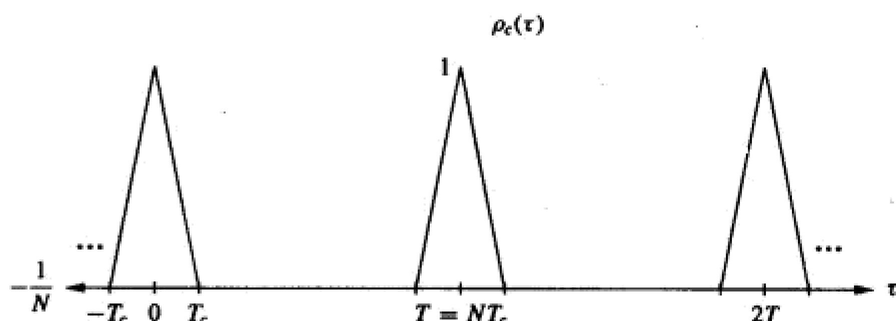


图13-8 周期为$T=NT_c$的自相关函数

周期性的自相关函数使ISI抑制问题变得复杂化。根据（13-16），对ISI的衰减量等于码元周期T_s上的自相关函数$\rho_c(\tau)$。因此，如果一个码元周期内有$N=T_s/T_c$个码片，那么解调器的相关就是在整个扩频码周期上进行，此时的$\rho_c(\tau)$就是式（13-19）。$N=T_s/T_c$的扩频码也叫*短扩频码*（short spreding code），此时图13-8所示的自相关函数的周期是$T=T_s$。如果多径时延是码元间隔的整数倍，则采用短码时前几个码元对当前码元的干扰将十分严重。加大扩频码的周期使$N>>T_s/T_c$，则只有时延非常大的才不能被抑制，不过时延很长时，路径损耗也很大，因此这些路径的功率很低。当$N>>T_s/T_c$时也叫个*长扩频码*（long spreding code）。长扩频码情况下，（13-17）的积分范围只是扩频码周期的一小部分：$T=T_s<<NT_c$，这样积分结果不再是（13-19），时延大于一码片时，衰减系数也不再是$-1/N$。这个局部时间上的周期自相关函数与其起点的位置有关，这使相关分析

变得很困难。对所有起点平均后的结果表明，对ISI的衰减程度约为$1/G$，处理增益近似等于每个码元中的码片数，即$G\approx T_sT_c$（见文献[6]的9.2节）。

虽然最大线性码在抑制ISI方面很好的性能，但它的有些特性使其在扩频多用户通信中不是最优的。给定码长N，最大线性码的个数较少，限制了多用户DSSS系统中同时共享系统带宽的用户数。有些m序列之间互相关性相对较差，归一化互相关值高达0.37（见文献[6]的9.2节）。所以，多用户扩频系统经常用互相关性更好的Gold码、Kasami码和Walsh码代替最大线性码，但是这些码在抵抗ISI方面不如最大线性码。13.4.1节将详细讨论这些码。

416

例13.3：某扩频系统采用的m序列周期为$T=T_s$，每个符号中有$N=100$个码片。假设同步器以$0.5T_c$的时延同步到直射路径。请问此时延偏差使信号降低了多少？

解：$\tau=0.5T_c$和$N=100$时，（13-19）给出的自相关函数$\rho_c(\tau)$为

$$1-\frac{|\tau|(1+1/N)}{T_c}=1-\frac{0.5T_c(1+1/100)}{T_c}=1-0.5\times1.01=0.495$$

信号部分乘上$\rho_c(\tau)$后，功率降低了$\rho_c^2(\tau)=0.495^2=0.245=-6.11\text{dB}$。可见信号功率被严重衰减，说明了精确同步的重要性。下一节我们讨论同步问题。

13.2.3 同步

现在我们考虑图13-5中的同步部分。载波提取环路是单独的，故此假设解调中的载波和接收到的载波相位一致。同步器需将接收端产生的扩频码与通过信道达接收端的某个多径分量中的扩频码时间对齐。常用的方法是采用图13-9所示的反馈控制环路。其原理是调节整码生成器的时延τ，使$w(\tau)$达到最大值。下面将看到，在理想的情况下，$w(\tau)$达到最大时的扩频码与输入信号是完全同步的。

假设信道的冲激响应为$h(t)=\delta(t-\tau_0)$，其中τ_0是信道时延。忽略噪声，图13-5中同步器的输入为$z(t)=x(t-\tau_0)s_c(t-\tau_0)\cos(2\pi f_c(t-\tau_0))\cos(2\pi f_c t+\phi)$，$\phi$是载波恢复环路的相位。假设理想载波恢复，即$\cos(2\pi f_c(t-\tau_0))=\cos(2\pi f_c t+\phi)$，则$z(t)=x(t-\tau_0)s_c(t-\tau_0)\cos^2(2\pi f_c t+\phi)$。反馈环路在$\tau=\tau_0$时达到同步。假设$x(t)$是采用矩形脉冲的BPSK，则$s_l=\pm\sqrt{E_b},x(t)=\pm\sqrt{2E_b/T_b}$，在一个比特时间内是常数，再假设扩频信号采用$m$序列。后面将讨论一般扩频码和调制的情形。若扩频码周期为$T=NT_c$，则此周期内的自相关函数为（13-19），示于图13-7。令$\lambda=\sqrt{2/T_b}$，则根据（13-18）有

$$\begin{aligned}w(\tau)&=\frac{\lambda}{T}\int_0^T s_l s_c(t-\tau_0)s_c(t-\tau)\cos^2(2\pi f_c t+\phi)\,dt\\&\approx\frac{0.5\lambda s_l}{T}\int_0^T s_c(t-\tau_0)s_c(t-\tau)\,dt=0.5\lambda s_l\rho_c(\tau-\tau_0)\end{aligned}\tag{13-21}$$

由于$s_l=\pm\sqrt{E_b}$，并且在$\tau-\tau_0=0$时$\rho_c(\tau-\tau_0)$最大，因此反馈回路将不断调节τ以使$|w(\tau)|$达到最大。假设$|\tau-\tau_0|>T_c$，则由式（13-19），$\rho_c(\tau-\tau_0)=-1/N$，此时同步器工作在图13-7的自相关函数三角区之外。此时反馈回路将持续调节τ，一般是按T_c步长调，调

节足够多的步数后，可以使$|w(\tau)|$的值大于$1/N$。此时对应有$|\tau-\tau_0|<T_c$，同步器与理想同步的时间差在一个码片周期内，这一过程也称为**粗同步**（coarse synchronization）或**捕获**（acquisition）。通常信道中会有多个路径，此时粗同步将同步于高于门限的第一个径上。417

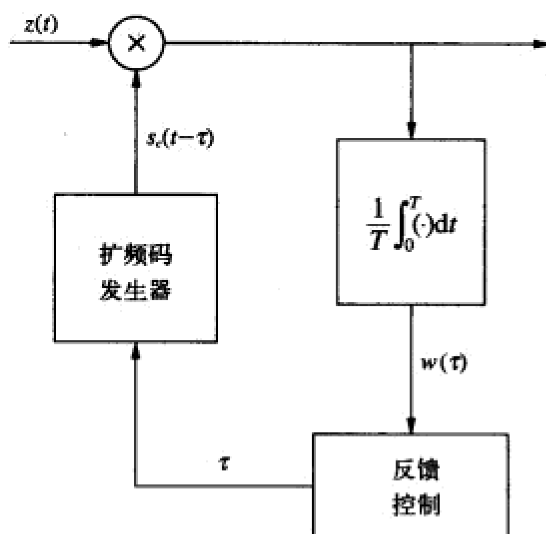


图13-9 DSSS中的同步环路

也可以按并行方式进行捕获，这样的系统中有多个并行的相关分支，分别用不同的时延与接收信号相关，各分支时延差是码片周期的整数倍，同步器将锁定到相关值最大的分支上。下一节讨论的RAKE接收机就采用了与此类似的结构将不同时延的多径信号进行同相合并。无论哪一种同步方法，粗同步一般用周期较短的短码以减少捕获时间。如果所用的是长码，可将反馈环路中的积分限制在扩频码周期的一小时间段内，从而缩短捕获时间。在这种情况下，只要局部自相关当时延大于一个码片周期时很小，当时延小于一个码片周期时高于某个门限值，就可以用这个相关结果来判断是否达到了粗同步。粗同步之后的微调一般采用长码在整个扩频码周期内积分，以使同步尽可能精确。418

一旦达到粗同步，反馈控制环路就微调τ值以使$\tau\approx\tau_0$，这个过程称为**细同步**（fine synchronization）或**跟踪**（tracking）。假如粗同步得到$\tau-\tau_0=T_c$，则同步器处在图13-7三角相关函数的最右边，随着τ减小，$\tau-\tau_0$趋向于0，三角形朝着峰值方向往回走。一旦达到峰值，同步器的时延就锁定在τ_0处。由于信道的时变性以及干扰、多径和噪声等的存在，τ必须要不断调整才能使达到动态最优的同步状态。扩频系统的同步跟踪采用的技术一般就是5.6.3节介绍的窄带系统中的定时提取技术。

采用其他扩频码字时的捕获和跟踪与此类似。所有周期扩频码的自相关函数都在零点处达到最大值，它们的粗同步和微同步都是不断调节时延以使自相关积分达到最大。同步性能与自相关函数的形状有密切的关系。尖锐的自相关函数能使同步更为精细。信道中的噪声、衰落、干扰和ISI会使图13-9中积分器的输出产生失真，这些因素会增大粗同步和细同步的复杂性。

当信号$s(t)$不是二进制，或者当码元周期小于扩频码周期时，积分器的输出将与积分区间内的具体数据符号有关。这种情形类似于5.6节中窄带系统在未知数据时的定时提取和载波提取，也可以采用类似的方法解决。前面的讨论中假设接收端的载波提取环路是理想的。扩频系统中的载波提取可以采用5.6节所述的技术，但必须要有所改动，因为扩频码对载波提取有一定的影响[7]。同步问题是扩频通信的一个关键问题，尤其在时变通信环境下更是一个难点。人们在同步技术的开发和分析方面已经开展了大量的工作，详见文献[2]的4.5节、文献[8]的第6章、文献[9]的第4部分中的1-2、文献[10]的12.5节。

13.2.4 RAKE接收机

图13-5中的扩频接收机只同步在接收信号的一个多径分量上，一般是粗同步时捕获的第一个高于门限值的分量。它的问题一是所同步的径有可能不是最强径，二是它把其他径当作了干扰。较为复杂的接收机有多个分支，每个分支同步于不同的多径分量，这种接收机称为RAKE接收机[1]。RAKE接收机假设每个码片周期倍数处都有一个多径分量，因此接收机中相邻支路的扩频码延时为T_c，如图13-10所示。每个多径分量的载波相位与其时延有关，各支路的相干解调器可以确定出这个相位。RAKE接收机实质上是一种分集合并，扩频后，大于一个码片周期的多径信号可以分辨，从而形成了路径分集。第7章所介绍的各种合并方法都可以用在RAKE接收机中，实际中最常用的是MRC。 419

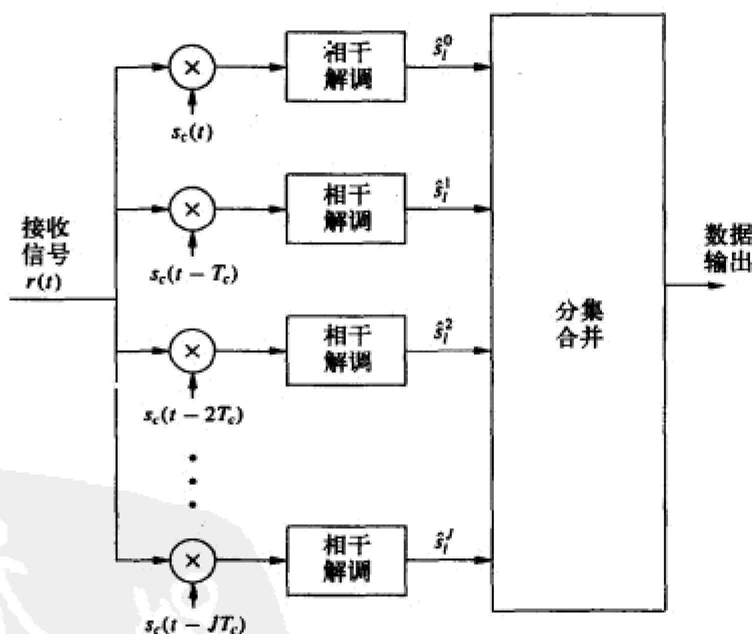


图13-10 RAKE接收机

1. RAKE（犁耙）这个名字是由其发明者Paul Green和Bob Price命名的。起这个名字是因为RAKE接收机的多支路结构很像一个犁耙，它也如犁耙一样通过多个分支犁起多个路径上的能量。RAKE接收机是1950年发明的，目的是为了解决高频越洋扩频链路所遇到的电离层多径问题。

为了便于研究RAKE接收机，假设信道冲激响应为$h(t)=\sum_{j=0}^{J}\alpha_j\delta(t-jT_c)$，$\alpha_j$是第$j$径的增益。只要适当建立$\alpha_j$的统计模型，那么3.4节所给出的这个模型基本能适用于各种多径环境。室外环境中α_j的经验统计特性见文献[11]。在这种模型下，图13-10中RAKE接收机中的不同支路将同步于不同时延的多径分量，并对其进行相干解调。J越大则复杂度也越大，但同时分集数也越大。根据式（13-14）和式（13-15），第i条分支的解调输出为：

420

$$\hat{s}_l^i=\alpha_i s_l+\sum_{\substack{j=1\\ i\neq j}}^{J-1}\alpha_j\rho_c(iT_c-jT_c)s_{lj}+n_l^i \tag{13-22}$$

其中s_l是在$[lT_s,(l+1)T_s]$内所发送的符号，也就是直射路径上的符号；s_{lj}是在$[lT_s-jT_c,(l+1)T_s-jT_c]$时间内发送的符号，假设这段时间内码元不变，$n_l^i$是第$i$支路中解调输出的噪声。如果在$[lT_s-jT_c,(l+1)T_s-jT_c]$内码元有变化，式（13-22）中的ISI项比较复杂，将涉及局部自相关函数。但不管码元是否有变化，对ISI的衰减都约为$\rho_c((i-j)T_c)$。图中的分集合并器将所有的解调输出合并在一起。选择合并输出的是路径增益α_i最大的那条支路的输出$\hat{s}_l^i$，等增益合并则将各支路的输出用相同的权重合并，最大比合并的权值则等于各支路的信噪比。如果考虑ISI的话，最大比合并的权值则是信干噪比（signal-to-interference-plus-noise power ratio，SINR）。如果$|\tau|>T_c$时，$\rho_c(\tau)\approx 0$，可以忽略各支路上的ISI，此时RAKE接收机的性能与J分支的分集技术相同。Turin在[9]中深入研究了RAKE接收机在经验信道模型下的性能。

由于扩频需要的带宽比其他分集技术大得多，所以一般不会纯因分集而选择扩频技术。如果是出于多用户通信或者干扰抑制等其他考虑选择了扩频，则RAKE接收机是一种获得分集增益的简单方法。

13.3 跳频扩频系统（FHSS）

端到端的跳频扩频系统如图13-11所示。频率合成器用扩频码产生跳频载波$c(t)=\cos(2\pi f_i+\theta_i(t))$，再将信号调制到这个载波频率上。调制方式可以是相干、非相关调制或差分相干。不过由于宽带跳频时维持相干的载波相位比较困难，所以非相干调制较为普遍（见文献[9]的第2部分）。在接收端，同步器用来产生与接收信号同步的本地扩频码。一旦建立同步后，频率合成器中用扩频码产生跳频载波，再用它进行解调。如果采用的是非相干或差分相干调制，则不必使接收端的载波与发送端同步。

421

和DSSS一样，FH系统的同步过程一般也分为两个过程。首先用粗同步使接收跳频序列与发送的调频信号同步在一个跳频间隔T_c内。粗同步的过程与DSSS类似：将叠加了噪声的跳频信号与接收端的跳频序列相乘，计算出乘积的能量，如果能量超过某个门限值则表明已经达到粗同步，否则就将接收到的跳频信号延迟T_c，再重复这个过程。粗同步也可以用多个跳频序列并行完成，各个序列之间的时间差是T_c的不同整数倍。完成粗同步后，就不断地微调跳频器的定时，以使本地跳频序列与接收信号的相关值最

大。更多关于FH同步及同步误差对系统性能的影响请参见文献[9]的第4部分。

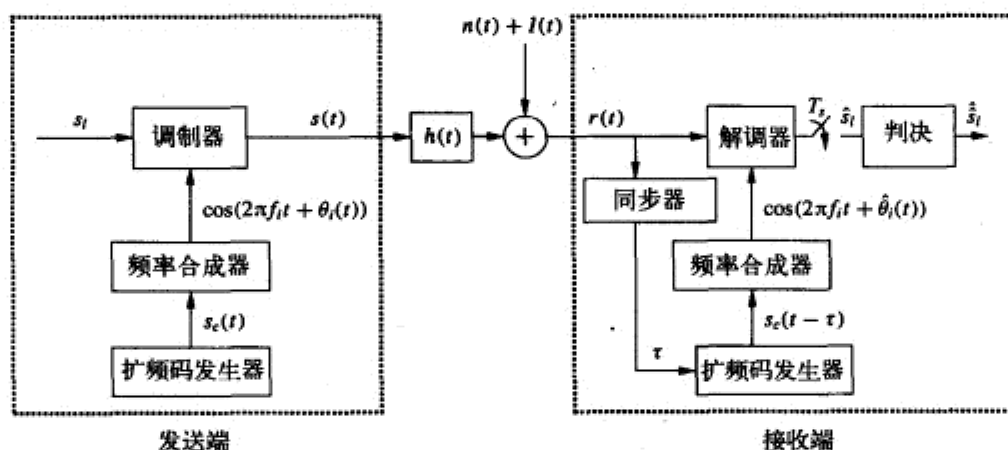


图13-11 FHSS系统模型

本章13.1节已经讨论过多径对FH系统的影响。如果多径时延大于跳频间隔，系统中不存在衰落。当多径能造成平衰落或频率选择性衰落时，系统性能的分析与慢变非跳频系统相同。不过窄带干扰对FH系统误码率的影响难以分析。窄带干扰对误码率的影响与干扰信号的具体特性以及干扰对具体调制的影响有关，下面讨论这一点。

我们来考虑无编码的SFH系统的误码率，慢跳频系统中可假设干扰在一个码元周期内是固定的。快跳频上问题要复杂得多，因为干扰在码元周期内是变化的，使得我们难以得出干扰的统计特性及其对误码率的影响。假设SFH系统的N个频带中有M个存在窄带干扰，若发送信号在整个频带中均匀跳频，则遇到干扰的概率是M/N。于是误码率为：

$$\begin{aligned}P_s &= p(\text{symbol error} \mid \text{no interference})p(\text{no interferenc}\\&\quad + p(\text{symbol error} \mid \text{interference})p(\text{interference})\\&= \frac{N-M}{N}p(\text{symbol error} \mid \text{no interference})\\&\quad + \frac{M}{N}p(\text{symbol error} \mid \text{interference})\end{aligned} \tag{13-23}$$

422

不存在干扰时的误码率就是信噪比为γ_s的AWGN信道的误码率，记其为P_s^{AWGN}。注意γ_s是无干扰时解调器输入端的信噪比，所以解扩过程对多径的消除不影响这个信噪比的数值，但它与所在频率上留下的多径分量的信道增益有关。对于大多数相关调制有 $P_s^{\text{AWGN}}\approx\alpha_M Q\left(\sqrt{\beta_M\gamma_s}\right)$，其中$\alpha_M$和$\beta_M$取决于具体的调制方式（见6.1.6节）。非相干调制和差分相干调制的P_s^{AWGN}一般很复杂（见文献[12]的1.1节）。给定P_s^{AWGN}后，需要确定的是有干扰时的误码率p（符号错误|有干扰），记为P_s^{INT}。则式（13-23）成为：

$$P_s=\frac{N-M}{N}P_s^{\text{AWGN}}+\frac{M}{N}P_s^{\text{INT}} \tag{13-24}$$

P_s^{INT} 与干扰特性有关。如果干扰的特性在信号带宽内如同加性高斯白噪声，这样的干扰称为部分频带噪声干扰（partial band noise jammer）。此时可将干扰视为一个功率为N_J 的加性高斯白噪声，于是总噪声功率为N_0B+N_J，信干噪比为

$$\gamma_s^{\text{INT}} = \gamma_s \frac{N_0 B}{N_0 B + N_J}$$

因此

$$P_s^{\text{INT}} = P_s^{\text{AWGN}}(\gamma_s^{\text{INT}}) \tag{13-25}$$

假设干扰是一个有一定相移的单音，则图13-11中的解调输出 $\hat{s}_l$ 为：

$$\hat{s}_l = a_l s_l + n_l + I_l \tag{13-26}$$

其中α_l 是解扩后接收信号的信道增益，n_l 是高斯白噪声， $I_l = \sqrt{I}e^{j\phi_l}$ 是相移为ϕ_l 的干扰项。注意宽带衰落是频率选择性的，因此：(i) 信道增益α_l 与载波频率有关，(ii) 有可能跳到信道增益很差的频率上。干扰项I_l 的影响与调制方式有关。例如对于相干MPSK，设 $\angle s_l = 0$，则

$$P_s = 1 - p(|\angle(a_l s_l + n_l + I_l)| \leqslant \pi/M) \tag{13-27}$$

一般来说，无论是相干调制还是非相干调制下，计算P_s 都需要得到随机相位 $\angle(n_l + I_l)$ 的分布。文献[9]的第2部分～第3部分给出了这个分布，并在不同干扰模型下，给出了非相干、相干和差分相干调制的P_s。跳频系统经常采用编码或交织编码来对抗频率选择性衰落和窄带干扰，文献[9]的第2部分对此进行了分析。

423

13.4　多用户DSSS系统

利用扩频码的特性，扩频通信可以使多个用户共享相同的频谱，也称为扩频多址（spread-spectrum multiple access，SSMA），它是码分多址（code-division multiple access，CDMA）的一种。多用户扩频系统给每个用户分配一个唯一的扩频码字或跳频模式，用它来调制数据信号。扩频码或跳频模式可以是正交的，此时，若传播条件理想的话，用户之间没有干扰；也可以是非正交的，此时用户之间有干扰，但干扰能被扩频码降低。因此，虽然单用户扩频系统因其占用的带宽超过所需，从而频谱效率不高，但在相同频带下，多用户扩频系统所能支持的用户数能大于等于时分多址、频分多址等其他共享带宽的方式。不过，如果扩频是非正交的，用户之间就会有干扰。非正交可能是设计原因，也可能是信道失真所致。当用户间干扰很大时，所有用户的性能都会恶化。第14章和第15章将比较多用户系统及蜂窝系统中不同频谱共享方法的频谱效率。

多用户扩频系统性能还与它是下行（downlink）信道（一个发射机对多个接收机）还是上行（uplink）信道（多个发射机对一个接收端）有关。上行及下行信道如图13-12所示。下行信道也称为广播信道或前向链路，上行信道也称为多址信道或反向链路。DSSS系统的上下行性能有差异是因为：下行的所有发送信号一般都是同步的，因为它

PDG

们都来自同一个发射机，并且期望信号和干扰信号经过相同的信道到达期望信号的接收机；而上行信道一般是异步的，因为它们来自不同位置上的发射机，并且各用户发送的信号经过不同的信道到达接收机。本节分析多用户DSSS系统的上行和下行链路，13.5节分析多用户FHSS系统。 424

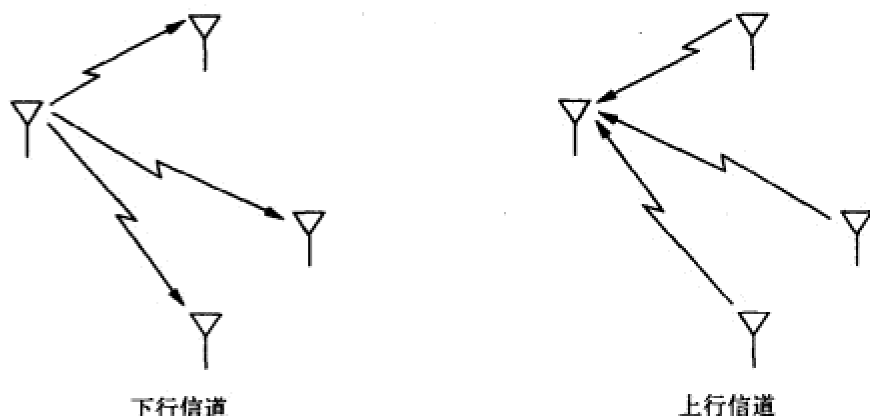


图13-12 下行和上行信道

13.4.1 多用户DSSS系统中的扩频码

多用户DSSS系统为每个用户分配一个唯一的扩频序列$s_{c_i}(t)$。扩频码的自相关决定了它的抗多径特性，不同扩频码之间的互相关决定了用户之间的干扰特性。异步用户i和j的信号到达接收机时有任意时延τ，它们的扩频码在一个码元周期内的互相关函数为

$$\rho_{ij}(t,t-\tau)=\frac{1}{T_s}\int_0^{T_s} s_{c_i}(t)s_{c_j}(t-\tau)\,\mathrm{d}t=\frac{1}{N}\sum_{n=1}^{N}s_{c_i}(nT_c)s_{c_j}(nT_c-\tau) \tag{13-28}$$

式中N是每符号的码片数。如果$s_{c_i}(t)$和$s_{c_j}(t)$的周期都是T_s，则互相关也如自相关那样只是时间差τ的函数，即$\rho_{ij}(t,t-\tau)\triangleq\rho_{ij}(\tau)$。同步用户的信号同时到达接收机，故$\tau=0$，互相关函数为

$$\rho_{ij}(0)=\frac{1}{T_s}\int_0^{T_s} s_{c_i}(t)s_{c_j}(t)\,\mathrm{d}t=\frac{1}{N}\sum_{n=1}^{N}s_{c_i}(nT_c)s_{c_j}(nT_c) \tag{13-29}$$

扩频码的互相关特性决定了用户间干扰的程度，理想情形是：对于异步用户有$\rho_{ij}(\tau)=0,\forall\tau,i\neq j$，对于同步用户有$\rho_{ij}(0)=0,i\neq j$，这样就可以完全消除用户间的干扰。满足这种互相关特性的扩频码叫正交码，否则是非正交码。对于异步用户，不可能在任意情况下都达到正交。对于同步用户，给定带宽时的正交码字的个数也是有限的。因此，同步多用户DSSS系统中的正交性要求会限制了可用的码字数，用户数也受到限制。下面介绍多用户DSSS系统中常用的扩频码。

Gold码

Gold码的自相关比m序列差，但设计合理时它有更好的互相关特性。Gold序列是两个长度为2^n-1的m序列的模二和，它继承了m序列的均衡性、游程特性和移位特性，因而也是伪随机序列。将两个时延分别为τ_1、τ_2的m序列模二加得到的序列对于不同的τ_1、τ_2是同的，Gold码正是利用这一点得到了大量不同的序列，从而可以允许多用户系统能容纳大量的用户。不过，随意选取一对m序列进行模二加所得到的序列可能有很差的互相关特性，因此Gold序列一般采用一对优选的m序列，通过优选使所得到的Gold序列有很好的互相关特性。Gold序列的自相关函数和通常的m序列不同。为了限制自相关值和互相关值，Gold在文献[13]中给出了优选m序列的方法，也见于文献[6]的9.2节，文献[8]的附录7及文献[14]。由m序列优选对构成的Gold序列的周期互相关是一个三值函数：

425

$$\rho_{ij}(\tau)=\begin{cases}-1/N\\-t(n)/N\\(t(n)-2)/N\end{cases}\tag{13-30}$$

其中

$$t(n)=\begin{cases}2^{(n+1)/2}+1 & n\text{奇}\\2^{(n+2)/2}+1 & n\text{偶}\end{cases}\tag{13-31}$$

它的自相关函数也是这三个值。

Kasami序列

Kasami序列也是由m序列产生的，其特性与优选对产生的Gold序列类似，但Kasami码的互相关比Gold码好。Kasami码分大集合和小集合两种。小集合Kasami码的产生方法是，先产生一个周期为2^n-1的m序列a，其中n为偶数，对序列a每隔$2^{n/2}+1$个元素进行取样，得到一个短序列a'，其周期为$2^{n/2}-1$，将序列a与a'的不同循环移位进行模二加就得到了小集合的Kasami序列。a'有$2^{n/2}-2$个不同的循环移位，再加上不移位的模二和$a+a'$以及原来的a，总共有$2^{n/2}$个周期均为2^n-1的序列。Kasami码的周期自相关及互相关也和Gold码一样是三值的，这三个值是

$$\rho_{ij}(\tau)=\begin{cases}-1/N\\-s(n)/N\\(s(n)-2)/N\end{cases}\tag{13-32}$$

式中$s(n)=2^{n/2}+1$。因为$|s(n)|<|t(n)|$，可见Kasami码比Gold码有更好的自相关和互相关特性。实际上，Kasami码达到了Welch下界，这个下界是任何$2^{n/2}$个周期为2^n-1的码所能达到的最小互相关和自相关，因此在这个意义下Kasami码是最优的（见文献[9]的第1部分第5章及文献[14]）。

426

大集合Kasima序列的产生方法与此类似，它的序列数比小集合更多，能使多用户系统容纳更多的用户，但大集合Kasami码的自相关和互相关要比小集合码差。为了产生大集合码，先产生一个周期为$N=2^n-1$的m序列a，n为偶数。对a每隔$2^{n/2}+1$个元素取样得到a'序列，每隔$2^{(n+2)/2}+1$个元素取样得到a''序列。然后将a序列分别与a'和a''的循环移位

序列进行模二加，这样就得到了大集合Kasami序列。当n是4的倍数时，大集合的序列数为$2^{3n/2}$，当$\text{mod}_4(n)=2$时，序列数为$2^{3n/2}+2^{n/2}$。大集合Kasami码的周期互相关和自相关函数是五值的：

$$\rho(\tau)=\begin{cases} -1/N \\ (-1\pm 2^{n/2})/N \\ (-1\pm(2^{n/2}+1))/N \end{cases} \tag{13-33}$$

这些值比小集合的相关值大，可见大集合Kasami码的自相关和互相关特性不如小集合Kasami码。

例13.4：对于长度为$n=10$的大集合及小集合Kasami序列，分别求出序列数以及周期互相关的最大值。

解：小集合有$2^{n/2}=2^5=32$个序列。根据式（13-32），互相关的最大值为

$$\frac{1}{N}[2^{n/2}+1]=\frac{1}{2^{10}-1}[2^5+1]=0.032$$

对于大集合，因为$\text{mod}_4(10)=2$，所以序列数为$2^{3n/2}+2^{n/2}=2^{15}+2^{10}=33792$，比小集合大了三个数量级。互相关的最大值为

$$\frac{1}{N}[2^{n/2}+2]=\frac{1}{2^{10}-1}[2^5+2]=0.033$$

互相关略有增大，它换来了序列数的显著增大。

Walsh－Hadamard码

长度为$N=T_s/T_c$并且同步的Walsh-Hadamard码在一个码元周期内是相互正交的，任意两个码序列的互相关为零。对于采用Walsh-Hadamard码的同步用户，如果信道不会破坏正交性的话，接收端能以无干扰的方式将其分离。（有延迟的多径分量和直射分量是不同步的，因此不同用户的多径分量会引起用户间干扰，它所造成的性能损失可以正交因子来衡量[15]）。下行的信号来自同一个发射机，可以做到同步，但上行链路中不同的用户位置不同，实现同步有一定的难度。因此DSSS系统的上行链路很少用Walsh-Hadamard码。长度为N的Walsh-Hadamard码通过$N\times N$的Hadamard矩阵$\boldsymbol{H}_N$产生。$N=2$的Hadamard矩阵为 427

$$\boldsymbol{H}_2=\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}$$

更大的Hamdamard阵用$\boldsymbol{H}_2$递归得到

$$\boldsymbol{H}_{2N}=\begin{bmatrix}\boldsymbol{H}_N & \boldsymbol{H}_N\\ \boldsymbol{H}_N & -\boldsymbol{H}_N\end{bmatrix}$$

$\boldsymbol{H}_N$的每一行是一个码片序列，所以Walsh-Hadamard码的序列数是N。因此采用Walsh-

Hadamard码的DSSS系统最多只能支持$N=T_s/T_c$个用户。DSSS占用的带宽近似为信息信号带宽的N倍，若将系统带宽分为N个不重叠的信道（频分），所能支持的用户数是一样的。同样的，若将时间分成N个正交时隙（时分），每个用户在其时隙内占用整个带宽，可支持的用户数也相同。因此，所有通过正交化信道使用户之间无干扰的多用户系统所能支持的用户数是基本相同的。

DSSS多用户系统的性能与扩频码的特性和传输信道有关。下面两节分别研究多用户DSSS系统在下行链路和上行链路中的性能。

13.4.2 下行信道

DSSS下行链路的发送端如图13-13所示，信道及接收端如图13-14所示。我们主要分析信号的同相分量，正交分量可类似分析。在下行链路中，发送端（基站）一般是同时发送所有用户的信号，接收端只是解调出用户自己的信号。因此可假设发射信号是同步的，于是可以采用Walsh-Hadamard码等正交的扩频码。由于使用正交码会限制用户数，所以下行链路并非一定会用正交码。

428

假设发送端向K个独立的用户发送信息。第k个用户的基带信号为

$$x_k(t)=\sum_l s_{kl}g(t-lT_s) \tag{13-34}$$

其中 $g(t)=\sqrt{2/T_s}$ 是矩形的成形脉冲，T_s为码元间隔，s_{kl} 是第k个用户在第l个码元间隔内发送的符号。发送端有K个支路，第k个支路将第k个用户的信号$x_k(t)$与扩频码 $s_{c_k}(t)$ 相乘，各个支路相加得到基带多用户信号为

$$z(t)=\sum_{k=1}^{K}x_k(t)s_{c_k}(t)=\sum_{k=1}^{K}\sqrt{\frac{2}{T_s}}s_{kl}s_{c_k}(t) \tag{13-35}$$

429

再同载波相乘得到所要发射的带通信号$s(t)$。

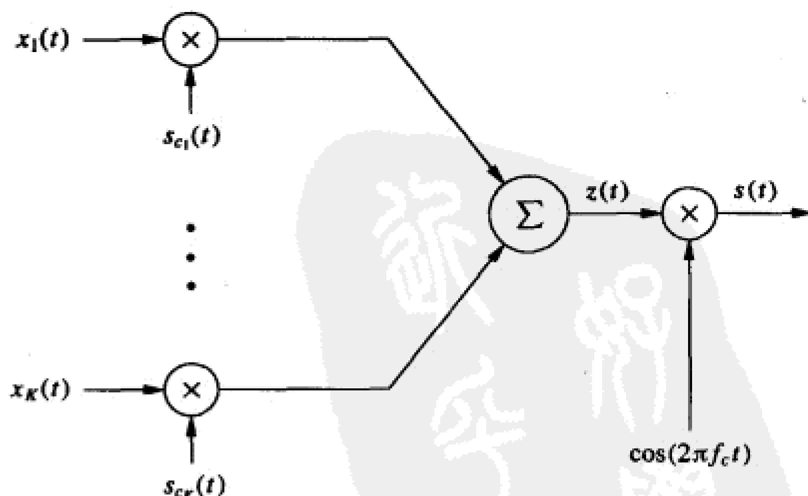


图13-13 下行链路的发送端

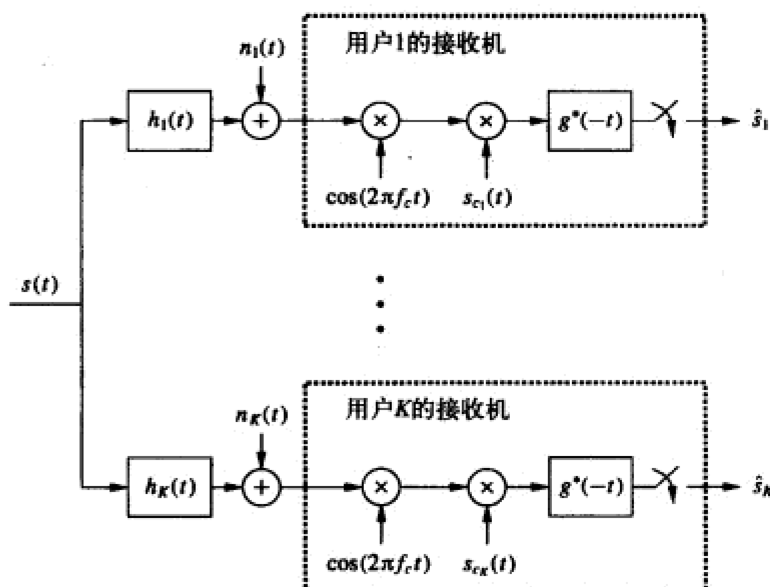


图13-14 下行信道及接收机

信号到达用户k时经过了冲激响应为$h_k(t)$的信道，并叠加了高斯白噪声，接收信号为$s(t)*h_k(t)+n(t)$。假设接收端有理想的载波恢复，第k个接收机的载波相应ϕ_k与接收信号理想匹配。因此我们将相移归一化为零，将时延归一化到直射径（或时延最小的径）信号。经过下变频后与k用户的扩频码$s_{c_k}(t)$相乘。假设$s_{c_k}(t)$与接收信号完全同步[1]。然后经过匹配滤波器解调（即乘以$\sqrt{2/T_s}$后在码元周期内积分）。再按T_s间隔取样得到相应码元周期内k用户发送符号的估计值。比较图13-5和图13-14可见，第k个用户的接收机就是一个单用户的匹配滤波检测。因此，如果没有多用户干扰，其性能也和单用户DSSS系统相同。但在存在多用户干扰时，解调输出将包括用户k的信号分量、其他用户的干扰以及噪声。在第l个码元间隔内的第k个用户的解调输出为

$$
\begin{aligned}
\hat{s}_k &= \sqrt{\frac{2}{T_s}}\int_0^{T_s}[s(t)*h_k(t)+n_k(t)]s_{c_k}(t)\cos(2\pi f_c t)\mathrm{d}t \\
&= \sqrt{\frac{2}{T_s}}\int_0^{T_s}[z(t)*h_k^{\mathrm{LP}}(t)]s_{c_k}(t)\cos^2(2\pi f_c t)\mathrm{d}t+\sqrt{\frac{2}{T_s}}\int_0^{T_s}n_k(t)s_{ck}(t)\cos(2\pi f_c t)\mathrm{d}t \\
&= \frac{2}{T_s}\int_0^{T_s}\left[\sum_{j=1}^{K}s_{jl}s_{c_j}(t)*h_k^{\mathrm{LP}}(t)\right]s_{c_k}(t)\cos^2(2\pi f_c t)\mathrm{d}t \\
&\quad+\sqrt{\frac{2}{T_s}}\int_0^{T_s}n_k(t)s_{c_k}(t)\cos(2\pi f_c t)\mathrm{d}t
\end{aligned}
$$

1. 由于存在多个扩频信号，多用户系统中的同步比单用户系统的同步更复杂。为此，某些系统中用一个扩频码的不同时移来产生不同用户的扩频扩频码。这些系统需要有专门的控制信道来通知接收机期望信号的时延是多少，详见文献[8]的第6章。

$$
\begin{aligned}
&= \frac{2}{T_s}\int_0^{T_s}[s_{kl}s_{c_k}(t)*h_k^{\mathrm{LP}}(t)]s_{c_k}(t)\cos^2(2\pi f_c t)\mathrm{d}t \\
&+\frac{2}{T_s}\int_0^{T_s}\left[\sum_{\substack{j=1\\ j\neq k}}^{K}s_{jl}s_{c_j}(t)*h_k^{\mathrm{LP}}(t)\right]s_{c_k}(t)\cos^2(2\pi f_c t)\mathrm{d}t \\
&+\sqrt{\frac{2}{T_s}}\int_0^{T_s}n_k(t)s_{c_k}(t)\cos(2\pi f_c t)\mathrm{d}t
\end{aligned} \tag{13-36}
$$

式中 $h_k^{LP}(t)$ 是 $h_k(t)$ 的等效基带响应的同相分量，s_{kl} 是所要恢复的第l个码元周期内的第k个用户的发送符号，s_{jl} 是第j个用户在这个码元间隔内的符号，它构成了干扰。式（13-36）中有三项，第一项只包含用户k的接收信号，第二项是其他用户产生的干扰，第三项是高斯白噪声，记噪声为n_k。在13.2节单用户系统的分析中我们已经得到了第一项和噪声项。第二项与信道 $h_k^{LP}(t)$ 和扩频码的特性有关。

430

为了了解多用户干扰的特性，先假设用户k的信道增益α_k为实值，且时延归一化为零，即$h(k)=\alpha_k\delta(t)$。于是，式（13-36）成为

$$
\begin{aligned}
\hat{s}_k &= \frac{2}{T_s}\int_0^{T_s}\alpha_k s_{kl}s_{c_k}^2(t)\cos^2(2\pi f_c t)\,\mathrm{d}t \\
&\quad+\frac{2}{T_s}\int_0^{T_s}\sum_{\substack{j=1\\ j\neq k}}^{K}\alpha_k s_{jl}s_{c_j}(t)s_{c_k}(t)\cos^2(2\pi f_c t)\,\mathrm{d}t+n_k \\
&\approx \alpha_k s_{kl}+\alpha_k\sum_{\substack{j=1\\ j\neq k}}^{K}s_{jl}\rho_{jk}(0)+n_k
\end{aligned} \tag{13-37}
$$

式中的$\rho_{jk}(0)$就是 $s_{c_k}(t)$ 和 $s_{c_j}(t)$ 的互相关函数，式中已假设所有用户同步[1]。定义

$$
I_{kl}=\alpha_k\sum_{\substack{j=1\\ j\neq k}}^{K}s_{jl}\rho_{jk}(0) \tag{13-38}
$$

I_{kl} 即为用户k解调输出端的多用户干扰。从式（13-37）中可见，用户k的码元s_{kl} 受到信道增益的衰减，但扩频解扩对它没有影响，这一点和单用户非扩频系统完全一样。噪声项n_k 也与和单用户非扩频系统一样。其他用户的干扰受到用户k的信道增益α_k 和扩频码互相关$\rho_{jk}(0)$的衰减。采用正交码（如Walsh-Hadamard）时，$\rho_{jk}(0)=0$，用户间没有干扰，采用非正交扩频码时，$\rho_{jk}(0)$与具体分配给用户j和用户k的扩频码有关，例如Gold码的$\rho_{jk}(0)$有三种可能的取值。注意，期望信号和干扰的传播路径相同，所以它们都受到

1. 如果用户是异步的（这在下行信道中非常少见），则（13-37）中的$\rho_{jk}(0)$应换为$\rho_{jk}(\tau_{jk})$，τ_{jk}是用户j与用户k的相对时延。这将假设s_{jl} 在积分区间内不变，不然的话，干扰项与积分区间内的符号值有关。

相同信道增益α_k的衰减，这一点与后面要讨论的DSSS系统上行链路不同。 431

当多用户系统中的干扰近似于高斯特性时，可将干扰项看成是额外的噪声，按SINR来得到系统的性能。不过，即使在干扰数较多的时候，高斯近似也不很准确[16]。所有干扰都通过共同的衰落α_k成为相关的，故此只能似为条件高斯。当干扰数足够大时，由中心极限定理可知这种近似是非常准确的[1]。用户k的SINR定义为解调输出端此用户的信号功率与多用户干扰和噪声的平均功率之比。将AWGN信道中的信噪比换成SINR就可以得到用户k误比特率。如果干扰功率远大于噪声功率，则可以忽略噪声，用SIR来分析性能。用户的SIR定义为该用户的信号功率与多用户干扰功率之比。噪声可以忽略的系统称为*干扰受限*（interference limited）的系统。无论是SINR还是SIR，平均干扰功率都与扩频序列和发送符号有关，其分析有可能非常复杂。因此，经常按随机序列来计算干扰的平均功率。此时，若系统有K个用户，每码元有N个码片，则SIR为（见文献[17]的2.3节）

$$\text{SIR} = \frac{N}{K-1} \approx \frac{G}{K-1} \tag{13-39}$$

式中$G \approx N$为系统的处理增益。注意这个结果与式（13-6）中随机序列任意干扰下的SIR一致。如果需要考虑噪声，可将噪声功率按码元能量E_s归一化后加到式（13-39）中，得到

$$\text{SINR} = \left(\frac{N_0}{E_s} + \frac{K-1}{G}\right)^{-1} \tag{13-40}$$

现在考虑更一般的信道$h_k(t)=\sum_{m=0}^{M-1}\alpha_{km}\delta(t-\tau_{km})$，其中$m$=0对应直射路径。假设理想的载波恢复，可将直射路径的时延τ_{k0}及相位归一化为零。解调输出仍然有三项：第一项是用户k的期望信号，第二项是其他用户的干扰，第三项是不受信道影响的高斯白噪声。有用信号部分的分析与13.2节对多径信道的分析相同，多径分量被扩频码的自相关函数衰减。多用户干扰的分析要比先前复杂得多。假设同步于用户k的直射路径，则解调输出端的多用户的干扰为

$$\begin{aligned} I_{kl} &= \frac{2}{T_s}\int_0^{T_s}\sum_{\substack{j=1\\ j\neq k}}^{K}\sum_{m=0}^{M-1}\alpha_{km}s_{j(l-l_{km})}s_{c_j}(t-\tau_{km})\cos(2\pi f_c(t-\tau_{km}))s_{c_k}(t)\cos(2\pi f_c t)\,dt \\ &\approx \sum_{\substack{j=1\\ j\neq k}}^{K}\sum_{m=0}^{M-1}\alpha_{km}s_{j(l-l_{km})}\cos(2\pi f_c\tau_{km})\rho_{jk}(\tau_{km}) \end{aligned} \tag{13-41}$$

432

式中$s_{j(l-l_{km})}$是用户j在第$(lT_s-\tau_{km})$个码元间隔内的发送符号，简单起见，式中已假设此码元间隔内符号不变。比较式（13-38）和式（13-41）可以看出，多径信道的影响有两点。首先是干扰项变多，原先是$K-1$项，现在是$(K-1)M$项，因此每个干扰用户都带来M项干扰项每个干扰用户的每一径构成一项。此外，即使用户同步，互相关也不再是按τ=0

1. 即便不是独立同分布，只要随机变量是不相关的，这一点也成立。

取，也就是说，多径破坏了信道的同步性。这一点对Walsh-Hamadard这样的正交码很重要，正交码一般只在时延为零时互相关才是零。因此在多径信道中，采用Walsh-Hamadard码的系统也有多用户干扰。对扩频码进行均衡可以消除用户间的干扰。

例13.5：考虑一个扩频因子为$N=B_s/B=100$的DSSS下行链路。假设系统是干扰受限的，且所有用户的信道都是单径的。若采用BPSK调制，要求每个用户的误比特率小于10^{-3}，那么此系统最多能容纳多少用户？

解：对于BPSK，$P_b=Q\left(\sqrt{2\gamma_b}\right)$，可求出$\gamma_b$=6.79dB时$P_b=10^{-3}$。因为是干扰受限系统，故可按SIR等于$\gamma_b$=6.79dB来求用户数$K$：

$$\mathrm{SIR}=\frac{N}{K-1}=\frac{100}{K-1}=10^{0.679}=4.775$$

解出$K\leqslant 1+100/4.77=21.96$。$K$只能是整数，而我们要求$P_b\leqslant 10^{-3}$，因此对21.96向下取整的到用户数为21。系统设计者有可能会接受轻微的误码率损失而建一个22用户的系统。

13.4.3 上行信道

现在考虑DSSS系统中的上行信道。多用户DSSS系统利用扩频码的特性分离不同用户的信号。上行和下行的主要区别在于，下行的信道对有用信号和干扰信号是相同的，而在上行链路中，不同用户的信号经历了不同的信道。这一点产生了*远近效应*（near-far effect），即上行链路中离接收端很近的用户会严重干扰远处的用户。

对于有K个用户的上行链路，每个用户的发送端和信道如图13-15所示。由于用户不在同一位置，所以发送信号一般是异步的。异步时的分析一般比同步复杂，性能也比同步时差。图13-15中用户k的基带信号是$x_k(t)$，我们也和下行信道时一样，只分析同相信号，并假设采用矩形脉冲。$x_k(t)$与扩频信号$s_{c_k}(t)$相乘后上变频，得到发射信号$s_k(t)$。（注意每个用户的载波相位不同）。这个信号通过用户k的信道传输，信道冲激响应是$h_k(t)$。所有用户的信号经过各自的信道后在接收端叠加，并叠加了高斯白噪声$n(t)$。

433

上行接收信号为

$$r(t)=\left[\sum_{k=1}^{K}(x_k(t)s_{c_k}(t)\cos(2\pi f_c t+\phi_k))*h_k(t)\right]+n(t) \tag{13-42}$$

接收机用K个分支来分别接收K个用户的信号，如图13-16所示。假设：(i) 用户k的直射路径的时延时为τ_k，(ii) 这个时延以及发射载波初相所形成的相位ϕ_k与接收端k支路恢复的载波相位ϕ_k'匹配。τ_k对同步用户是相等的。接收端的k支路将接收信号下变频，再乘以用户k的扩频码，该扩频码与用户k的接收信号同步。解扩后的信号经过匹配滤波器，再进行采样得到相应的发送符号估计值。比较图13-15和图13-16可见，上行接收机的一个支路是一个单用户DSSS系统的匹配滤检测器，整个上行接收机就是一组单用户的匹配滤波检测器。若无多用户干扰，各用户的性能与单用户系统相同。考虑多用户干扰时，

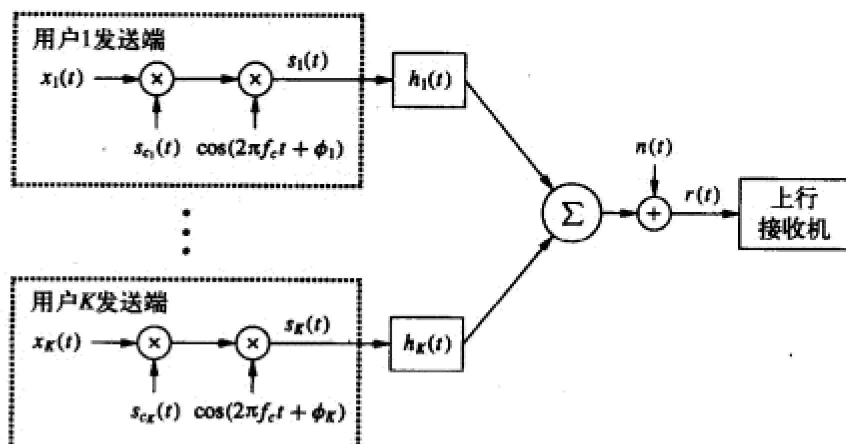


图 13-15

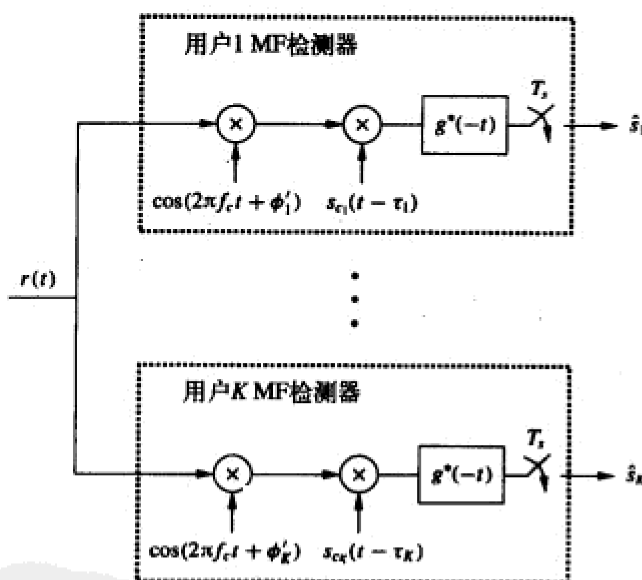


图 13-16

第l个码元间隔内k支路的解调输出为

$$\begin{aligned}\hat{s}_k &= \sqrt{\frac{2}{T_s}}\int_0^{T_s}\left[\sum_{j=1}^{K}(x_j(t)s_{c_j}(t))*h_j^{\mathrm{LP}}(t)\right]s_{c_k}(t-\tau_k)\cos(2\pi f_c t+\phi_k')\cos(2\pi f_c t+\phi_j')\,\mathrm{d}t \\ &\quad + n_k \\ &= \frac{2}{T_s}\int_0^{T_s}[s_{kl}s_{c_k}(t)*h_k^{\mathrm{LP}}(t)]s_{c_k}(t-\tau_k)\cos^2(2\pi f_c t+\phi_k')\,\mathrm{d}t \\ &\quad + \frac{2}{T_s}\int_0^{T_s}\left[\sum_{\substack{j=1\\ j\neq k}}^{K}s_{ljk}s_{c_j}(t)*h_j^{\mathrm{LP}}(t)\right]s_{c_k}(t-\tau_k)\cos(2\pi f_c t+\phi_k')\cos(2\pi f_c t+\phi_j')\,\mathrm{d}t \\ &\quad + n_k\end{aligned} \tag{13-43}$$

434

其中n_k是高斯白噪声，$h_j^{LP}(t)$是等效基带滤波器$h_j(t)$（j=1, L, K）的同相分量，其直射路径的时延归一化为零，s_{ljk}是时间$[lT_s-\tau_j+\tau_k,(l+1)T_s-\tau_j+\tau_k]$内用户$j$的发送符号，假设它在此时间内不变。如果有变化的话（即码元在lT_s时刻改变），ISI就会更复杂，将涉及局部自相关函数，但对ISI的衰减量大致不变。式（13-43）有三项：第一项是用户k的期望信号，最后一项是高斯白噪声，这两项与单用户系统相同。第二项是其他用户的干扰，其中用户j对用户k的干扰($j\neq k$)与用户j的等效基带信道$h_j^{LP}(t)$及其扩频码有关。

假设每个用户信道只有增益α_j和时延τ_j，即$h_j^{LP}(t)=\alpha_j\delta(t-\tau_j)$，则第$l$个码元间隔内$k$支路的解调输出为

$$\hat{s}_k=\alpha_k s_{kl}+I_{kl}+n_l \tag{13-44}$$

式中第一项和最后一项与单用户系统相同。将$h_j^{LP}(t)=\alpha_j\delta(t-\tau_j)$代入式（13-43）得到干扰项$I_{kl}$为

$$\begin{aligned}
I_{kl}&=\frac{2}{T_s}\int_0^{T_s}\left[\sum_{\substack{j=1\\j\neq k}}^{K}s_{ljk}s_{c_j}(t)*\alpha_j\delta(t-\tau_j)\right]s_{c_k}(t-\tau_k)\cos(2\pi f_c t+\phi_k')\cos(2\pi f_c t+\phi_j')\,\mathrm{d}t\\
&=\frac{1}{T_s}\int_0^{T_s}\left[\sum_{\substack{j=1\\j\neq k}}^{K}\alpha_j s_{ljk}s_{c_j}(t-\tau_j)\right]s_{c_k}(t-\tau_k)[\cos(\Delta\phi_{kj})+\cos(4\pi f_c t+\phi_k'+\phi_j')]\,\mathrm{d}t\\
&\approx\sum_{\substack{j=1\\j\neq k}}^{K}\alpha_j\cos(\Delta\phi_{kj})s_{ljk}\frac{1}{T_s}\int_0^{T_s}s_{c_j}(t-\tau_j)s_{c_k}(t-\tau_k)\,\mathrm{d}t\\
&=\sum_{\substack{j=1\\j\neq k}}^{K}\alpha_j\cos(\Delta\phi_{kj})s_{ljk}\rho_{jk}(\tau_j-\tau_k)
\end{aligned} \tag{13-45}$$

435

式中$\Delta\phi_{kj}=\phi_k'-\phi_j'$，其中的近似基于$f_c>>1/T_s$。从式（13-45）可见，上行的多用户干扰也和下行一样受到扩频码互相关的衰减。一般$\tau_j\neq\tau_k$，所以上行链路是异步的，因此上行链路一般不用要求同步接收的正交码（如Walsh-Hadamard码）。上行链路的另一个特点是，用户符号及多用户干扰所经历的信道衰减不同。用户k的信号受到α_k衰减，而来自用户j的干扰受到α_j的衰减。如果$\alpha_j>>\alpha_k$，即使经过扩频码的互相关衰减，干扰仍然会严重恶化系统的性能。

现在考虑干扰受限的上行信道。先假设刚开始所有用户到达接收端的功率相同。假设采用随机扩频码，每码元有N个码片，各用户的码序列起点随机、载波相位随机，则异步系统的平均SINR为[18]

$$\mathrm{SINR}=\left(\frac{K-1}{3N}+\frac{N_0}{E_s}\right)^{-1} \tag{13-46}$$

对于干扰受限系统，忽略噪声的影响，得到SIR为

$$\mathrm{SIR}=\frac{3N}{K-1}\approx\frac{3G}{K-1} \tag{13-47}$$

式中$G\approx N$是系统的处理增益。式（13-46）和式（13-47）称为SINR和SIR的标准高斯近似。对于任意的系统应用这两个式子时必须要注意的是，SIR和SINR与扩频码特性、定时、载波相位以及系统的其他特性有关。为了提高准确性，人们针对实际系统对标准高斯近似进行了修正，不过这些修正式一般比较复杂，而且对准确性的提高也不是非常大（见文献[6]的9.6节）。对于非随机的扩频码，式（13-47）可修正为

$$\mathrm{SIR}=\frac{3N}{\xi(K-1)}\approx\frac{3G}{\xi(K-1)} \tag{13-48}$$

436

式中ξ是反映扩频码互相关性的一个常数，它与扩频码的特性及其他的系统假设有关。标准高斯近似对应$\xi=1$。对于PN序列，视系统假设之不同，$\xi=2$[19]或$\xi=3$[20]。

现在假设所有$K-1$个干扰项的信道增益$\alpha>>\alpha_k$，则用户k的SIR为

$$\mathrm{SIR}(k)=\frac{\alpha_k^2 3N}{\alpha^2\xi(K-1)}\approx\frac{\alpha_k^2 3G}{\alpha^2\xi(K-1)}\ll\frac{3G}{\xi(K-1)} \tag{13-49}$$

信道增益的不同使用户k的SIR损失了α_k^2/α^2，这就是远近效应，用户远离接收机时，信道增益一般要比近处的干扰小很多。在衰落信道中，α_k是随机的，它能降低码字的互相关，从而增大SIR。衰落对平均互相关的影响可通过修正式（13-48）中的ξ来体现，具体数值与扩频码的特性、系统假设、衰落信道的特性有关[21]。

对于多径信道，$h_j^{LP}(t)=\sum_{m=1}^{M}\alpha_{jm}\delta(t-\tau_{jm})$，代入式（13-45）中可得到$k$支路上的多用户干扰为

$$I_{kl}\approx\sum_{\substack{j=1\\ j\neq k}}^{K}\sum_{m=1}^{M}\alpha_{jm}\cos(\Delta\phi_{jkm})s_{ljm}\rho_{jk}(\tau_{jm}-\tau_k) \tag{13-50}$$

其中假设k支路同步于信道延时τ_k，$\Delta\phi_{jkm}$是相对的相位差，s_{ljm}是用户j在时间$[lT_s-\tau_k+\tau_{jm},(l+1)T_s-\tau_k+\tau_{jm}]$内的发送符号，假设它在此时间内不变。这个干扰也会造成远近效应：如果有一径的信道增益显著大于用户k的信道增益，SIR将会恶化。

克服远近效应的一种方法是在DSSS的上行链路上采用基于信道反转的功率控制。它让第k个用户以功率P/α_k^2发送，这样接收功率是与路径损耗无关的P。此时的SIR是式（13-48）。这种功控的缺点是，在某些衰落信道中，实行信道反转需要很大的发送功率（例如瑞利衰落需要无限大的功率）。信道反转也会对其他系统或其他同频用户产生严重的干扰。比如会加剧蜂窝系统中的邻小区干扰。尽管有这些问题，IS-95标准中移动台到基站的连接采用了信道反转，不过它把反转后的发射功率限制为最大发射功率。

437

例13.6： 某个干扰受限的DSSS系统的上行链路的处理增益为$G=B_s/B=100$，采用BPSK调制。所有用户的信道都是单径的。假设用户k的接收功率比其他用户低6dB。若要求误码

率小于10^{-3}，请分别按标准高斯近似的随机码$\xi=1$以及$\xi=3$的PN序列，求系统所能容纳的用户数。

解：如同前例，$P_b=10^{-3}$需要SIR=γ_b=6.79dB=4.775。我们用这个SIR来求系统可支持的最大用户数K。这是一个异步系统，$\alpha_k^2/\alpha^2=0.251(-6\text{dB})$，所以：

$$\text{SIR}=\frac{\alpha_k^2 3N}{\alpha^2\xi(K-1)}=\frac{0.251\times300}{\xi(K-1)}=\frac{75.3}{\xi(K-1)}=4.775$$

解得$\xi=1$时$K\leqslant 1+75.3/4.77\xi=16.78$，$\xi=3$时$K\leqslant 6.26$，所以系统只能容纳6~16个用户，比前例中的下行链路小很多。原因是上行系统中用户是异步的并且存在远近效应。本例体现出系统容量计算对扩频码特性ξ的敏感性。SIR近似与$1/\xi K$成正比，因此给定SIR时，系统可容纳的用户数与$1/\xi$成正比。

13.4.4　多用户检测

我们不一定要把SSMA系统中的干扰当噪声处理。如果接收端已知干扰信号的扩频码，多址干扰（MAI）是可以消除的。同时检测所有用户消除了用户间的干扰，这样就可以提高系统的性能，或者在系统性能给定的条件下增加用户数。此外，同时检测所有用户时，远近效应是个有利因素，因为和所有用户信道相同的情形相比，信道增益强的用户更容易检测，这个检测结果可用于后续的干扰抵消。利用多用户干扰的结构进行信号检测称为多用户检测（multiuser detection，MUD）。下行链路中一般不用多用户检测技术，原因有这样几点：首先下行链路是同步的，只要信道不破坏扩频码的正交性，就可以用正交码来消除用户间的干扰；其次，用户接收机的复杂度和功耗一般都有限，很难加入复杂的多用户检测功能；最后，上行链路的接收端本来就要接收所有用户的信号，所以任何上行的接收机都是一种多用户检测，只是不一定好而已。而下行链路中用户k的接收机只需要检测用户k的信号。这些原因使得多用户检测主要用在DSSS系统的上行中。

Verdú最早开创了多用户检测技术，他在文献[22~23]中导出了异步DSSS上行信道的最佳联合检测。其推导假设各用户的信道是不同增益的AWGN信道。这种信道下的最佳检测选出这样一个K个用户的符号序列，它能使假设发送这个序列时所能得到的信号和实际收到的信号之间的均方误差最小。信道的异步性使得任何码元间隔内的最佳检测都需要处理全部的接收波形。因为各用户的符号在时间上不是对齐的，所以对任何时间间隔，必须考虑所有重叠出现的符号，对所有重叠的符号也必须考虑相关的时间区间，因此不可能在有限时间区间内使信号处理达到最佳。文献[23]中的异步最佳多用户检测先是K个单用户匹配滤波器，然后用维特比序列检测算法联合检测所有用户。维特比算法在二进制调制时有2^{K-1}个状态，其复杂度随2^K增长。

438

同步用户的最佳检测比较简单，只需考虑一个码元间隔，不需要序列检测。同步检测的复杂度随$2^K/K$增长，而序列检测是随2^K增长。考虑两用户的同步上行信道，信道k的增益为α_k，采用二进制调制。一个比特时间内的等效基带接收信号为

$$r(t)=\alpha_1 b_1 s_{c_1}(t)+\alpha_2 b_2 s_{c_2}(t)+n(t) \tag{13-51}$$

其中b_k是用户k在这个比特间隔内发送的比特。最佳（最大似然）检测器的输出$b^*=(b_1^*,b_2^*)$满足

$$\arg\min_{(b_1,b_2)}\left[-\frac{1}{2\sigma^2}\int_0^{T_b}[r(t)-\alpha_1 b_1 s_{c_1}(t)-\alpha_2 b_2 s_{c_2}(t)]^2\,dt\right] \tag{13-52}$$

其中σ^2是噪声功率。上式等价于找出(b_1^*,b_2^*)使下面的代价函数最大：

$$L(b_1,b_2)=\alpha_1 b_1 r_1+\alpha_2 b_2 r_2-\alpha_1\alpha_2 b_1 b_2 \rho_{12} \tag{13-53}$$

其中

$$r_k=\int_0^{T_b} r(t)s_{c_k}(t)\,dt \tag{13-54}$$

$$\rho_{jk}=\int_0^{T_b} s_{c_k}(t)s_{c_j}(t)\,dt \tag{13-55}$$

上述分析可直接推广到K个同步用户的情形。记$\boldsymbol{r}=(r_1,\cdots,r_K)^T$，矩阵表示为[24]

$$\boldsymbol{r}=\boldsymbol{RAb}+\boldsymbol{n} \tag{13-56}$$

其中$\boldsymbol{b}=(b_1,\cdots,b_K)^T$是给定比特间隔内$K$个用户的比特向量，$\boldsymbol{A}$是信道增益$\alpha_k$的$K\times K$对角阵，$\boldsymbol{R}$表示扩频码的$K\times K$互相关矩阵。最佳比特序列$\boldsymbol{b}^*$将使下面的代价函数最大

$$L(\boldsymbol{b})=2\boldsymbol{b}^T\boldsymbol{Ar}-\boldsymbol{b}^T\boldsymbol{ARAb} \tag{13-57}$$

假设用搜索树进行优化，则最大化式（13-57）的复杂度随2^K增长，与异步时相同。除了复杂性外，最佳检测还有一个缺点是需要已知信道增益α_k。

439

降低对最优的要求可以降低复杂度。人们已经提出了许多次优的多用户检测技术，这些技术能在性能、复杂度、对信道信息的要求方面进行不同的折中。次优多用户检测主要分为线性和非线性两大类。线性检测器对图13-16中匹配滤波器组的输出进行线性变换或滤波，其复杂度与用户数呈线性关系，显著低于最佳检测。常用的线性检测是解相关检测[25]和最小均方误差检测（MMSE）。解相关检测简单地对相关矩阵$\boldsymbol{R}$求逆，得到

$$\hat{\boldsymbol{b}}^*=\boldsymbol{R}^{-1}\boldsymbol{r}=\boldsymbol{R}^{-1}[\boldsymbol{RAb}+\boldsymbol{n}]=\boldsymbol{Ab}+\boldsymbol{R}^{-1}\boldsymbol{n} \tag{13-58}$$

在大多数情况下，逆矩阵是存在的。不考虑噪声时得到的比特序列是原序列乘以信道增益。除了简单外，它可以完全消除MAI，并且不需要已知信道增益。缺点是噪声乘以逆矩阵后会变大。因此解相关多用户检测器类似于11.14.1中介绍的迫零均衡。

MMSE检测器找到这样一个矩阵$\boldsymbol{D}$，用它乘以匹配滤波器组的输出$\boldsymbol{r}$得到的$\boldsymbol{Dr}$和发送比特序列$\boldsymbol{b}$的平均均方误差能达到最小，也即

$$\arg\min_{\boldsymbol{D}} E[(\boldsymbol{b}-\boldsymbol{Dr})^T(\boldsymbol{b}-\boldsymbol{Dr})] \tag{13-59}$$

最佳的$\boldsymbol{D}$为[17][24][26]

$$\boldsymbol{D}=(\boldsymbol{R}+0.5N_0\boldsymbol{I})^{-1} \tag{13-60}$$

无噪声时，MMSE检测和解相关检测是一样的。低信噪比时，MMSE检测性能更好，因为它较好地平衡了MAI消除和噪声增强，这与11.4.2节中提到的MMSE均衡类似。

440

非线性多用户检测一般要比线性多用户检测性能好，特别对于无编码或简单编码的系统[27]。常用的非线性多用户检测技术有多级检测、判决反馈检测和串行干扰抵消。多级检测的每一级有一个传统的匹配滤波器组，第n级利用第$(n-1)$级的判决输出来消除第n级输入端的MAI。多级检测可用于同步系统[28]和异步系统[29]。判决反馈检测的设计思想与判决反馈均衡相同，它包括一个前馈滤波和一个反馈滤波，其中前馈滤波是相关矩阵$\mathbf{R}$的Cholesky分解。判决反馈多用户检测器也可应用于同步系统[30]和异步系统[31]。这两种检测器都需要已知信道增益，另外，判决错误会引起错误传播。干扰抵消则是先检测出一个或者多个用户，再把他们对其他用户的MAI减掉[32]。干扰抵消可以是并行的[33, 34]或串行的。并行干扰抵消检测出所有的用户[35]，再进行抵消；串行干扰抵消一次抵消一个用户，再对待检测的用户抵消干扰。并行干扰抵消时延小，对判决错误敏感度低，但存在远近效应。当一些用户的接收功率远小于其他用户时，这种远近效应会严重影响并行干扰抵消的性能。用户的接收功率不相等时，串行干扰抵消要优于并行干扰抵[24]。实际上，串行干扰抵消理论上可以达到香农极限（见第14章），实际中也能接近了香农极限[36]。串行干扰抵消存在错误传播的问题，它会使系统的性能显著恶化，可通过功率控制部分消除这种影响[37]。

文献[17]对不同的多用户检测技术及其性能进行了深入全面的介绍。文献[24]和[32]有一些简短的讲解。文献[38]介绍了多用户检测和均衡器结合。文献[39]分析了不同用户有不同的数据速率的多速率CDMA中的多用户检测。文献[40]中介绍了盲多用户检测、空时多用户检测和Turbo多用户检测。文献[27]分析了不同检测技术的频谱利用率。

13.4.5 多载波CDMA

多载波CDMA（multicarrier CDMA，MC-CDMA）结合了OFDM和CDMA的长处。它能有效抑制ISI，并能使多个用户共享同一信道。图13-17是单用户多载波CDMA的基带框图。数据符号s_l在所有N个子信道上发送，第i个子信道的s_l与扩频码$s_c(t)$的第i个码片$c_i=\pm 1$相乘。除了扩频码相乘是在频域外，它与标准的扩频技术相同。频域扩频的数据$(s_lc_1, s_lc_2, \cdots, s_lc_N)$用标准的方法进行多载波调制：先进行IFFT、再并串变换、再经过D/A变换得到已调信号$s(t)$。$s(t)$的频谱$S(f)$如图13-17所示，各子信道的载频是$(f_1, f_2, \cdots, f_N)$。

假设MC-CDMA信号通过频率选择性信道发送，信道的高斯白噪声是$n(t)$，子信道i的增益为常数α_i。接收端用与发送端相反的方式恢复出各个子信道上的数据：先经过A/D转换、再串并变换、再进行FFT。然后将第i个子信道接收的符号与第i个码片c_i相乘，再乘上加权系数β_i，再将这些项相加得到最终的符号值$\hat{s}_l$。

在多用户MC-CDMA系统中，每个用户都用其唯一的扩频序列$s_{c_k}(t)$按图13-17所示进行调制。例如对于两用户系统，用户1的扩频码是$s_{c_1}(t)$，码片序列是$(c_1^1, \cdots, c_N^1)$，发射信号是$s_1(t)$；用户2的扩频码是$s_{c_2}(t)$，码片序列是$(c_1^2, \cdots, c_N^2)$，发射信号是$s_2(t)$。两个用户同时发射，它们的信号在空中相加，如图13-18所示，图中$s_l^1(s_l^2)$是用户1（2）在第l个码元间隔内的符号。用户之间的干扰会被扩频码的互相关抑制，如同单载波扩频系统。不过，MC-CDMA把每个用户的信号扩展到了多个独立衰落的子信道上，因此能获得频率分集增益，这使它比传统的扩频系统有更好的性能。

441

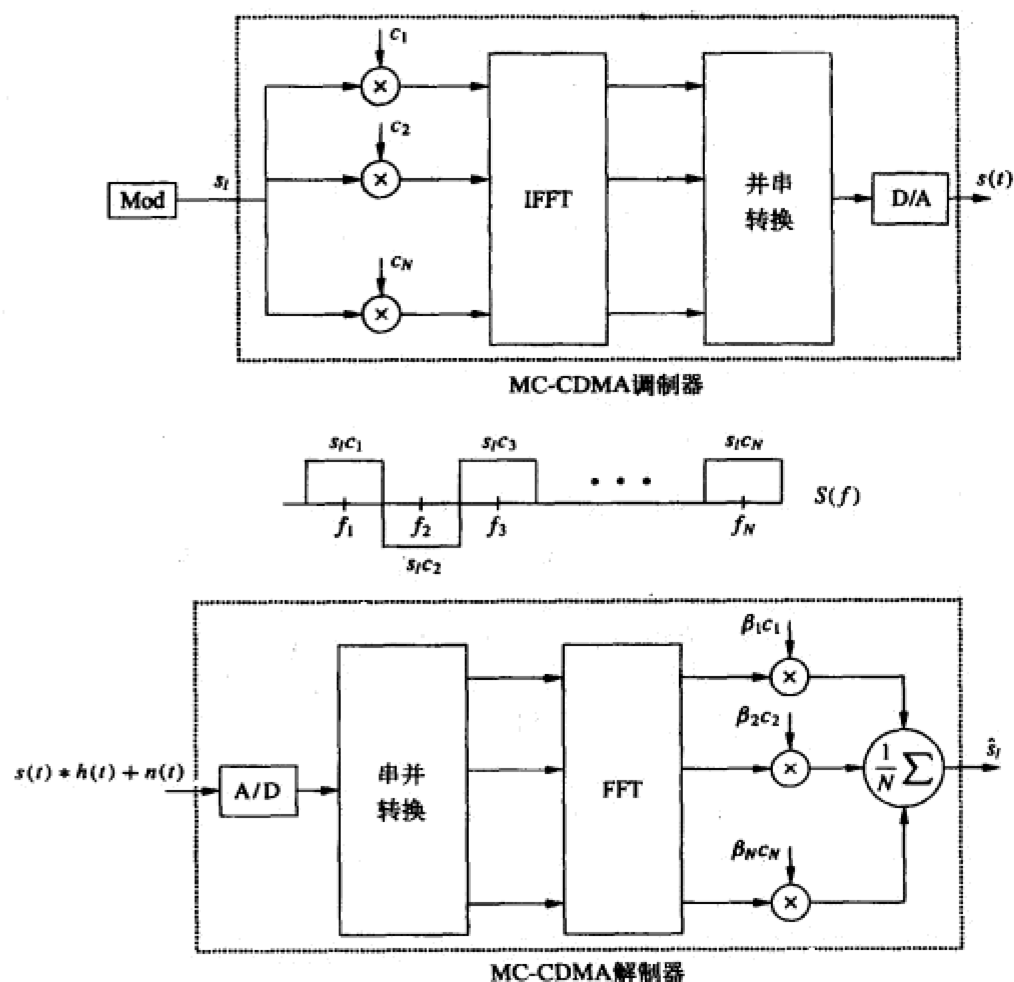


图13-17 多载波CDMA系统

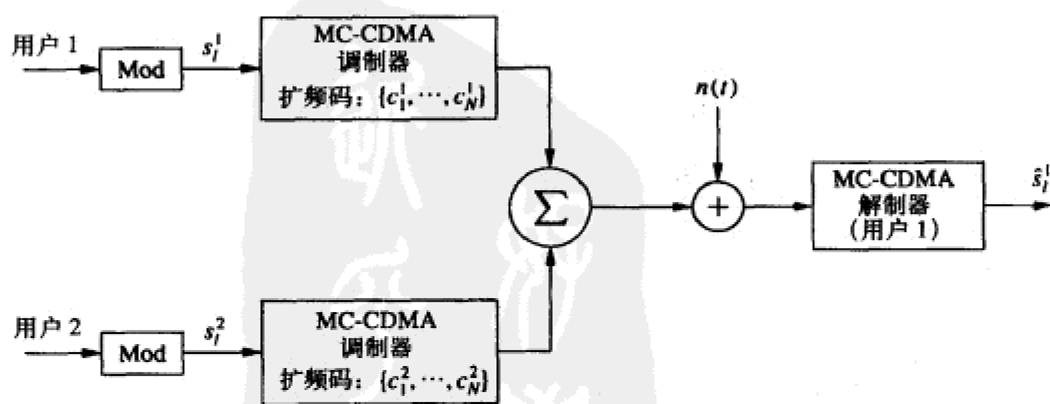


图13-18 两用户的多载波CDMA系统

442

13.5 多用户FHSS系统

多用户FHSS系统给每个用户分配一个唯一的扩频码 $s_{c_i}(t)$，用这个码来产生跳频模式。如果扩频码是正交的，并且用户在时间上是同步的，则不同的用户不会产生碰撞，此时的性能和单用户FH系统相同。但如果用户不同步或者扩频码不正交的话，多个用户能同时跳到同一个信道，从而发生碰撞。碰撞时发送的符号极有可能出错，因此多用户FHSS系统一般采用纠错编码来对抗碰撞。

多用户FHSS也称为FH-CDMA或FH-SSMA，主要用在上行链路中。由于FH具有抗干扰、不易被截获和检测的特性，所以多用户FHSS是军事通信的首选。FH-CDMA曾是第二代数字蜂窝系统中的一个候选提案（见文献[2]的9.4节），但未被采纳。文献[8]的第11章、文献[41~42]讨论了跳频多址和直序多址之间的折中问题。大部分研究认为，至少在异步情况下，跳频多址可支持的用户数不及直序扩频。和DS-CDMA相比，FH-CDMA能干扰到更远距离处的通信，因为它的信号频谱没有被扩展（见文献[8]的第11章）。FH相对于DS的最大优点是没有远近效应，它是通过频率跳跃来躲避干扰，不是通过处理增益来减弱干扰，因此能对抗强干扰引起的性能恶化。可以将跳频技术和其他多址接入技术结合以发挥各自的长处。例如GSM数字通信系统采用了慢跳频和时分多址的结合，跳频用来平均邻小区干扰。工作在2.4GHz开放频段的蓝牙也选择了FH，其目的是为了能用低成本、低功耗的非相干FSK。 443

习题

13.1 本题用随机扩频来推导式（13-6）的SIR。第i个接收支路的相关输出为式（13-5）：

$$x_i = \int_0^T x(t)s_i(t)\,dt = \sum_{j=1}^{N}(s_{ij}^2 + I_j s_{ij})$$

(a) 证明以发射信号$s_i(t)$为条件时，x_i的条件数学期望为$E[x_i|s_i(t)]=E_s$。

(b) 假设信号等概出现($p(s_i(t))=1/M$)，证明$E[x_i]=E_s/M$。

(c) 证明$\mathrm{Var}[x_i|s_i(t)]=E_s E_j/N$。

(d) 假设信号等概，证明$\mathrm{Var}[x_i]=E_s E_j/NM$。

(e) SIR定义为 $\mathrm{SIR}=\dfrac{E[x_i]^2}{\mathrm{Var}[x_i]}$，证明 $\mathrm{SIR}=\dfrac{E_s}{E_j}\times\dfrac{N}{M}$

13.2 画出两比特间隔$[0, 2T_b]$内DSSS信号$s(t)s_c(t)$的波形。假设$s(t)$是载波频率为100MHz的BPSK调制，比特周期$T_b=1\mu s$，两个数据比特是10，每比特对应10个扩频码片，码片序列是正负交替的±1，第一个码片是+1。

13.3 某FH系统中的信号经过两径信道传输。反射路径相对于直射路径的延时为$\tau=10\mu s$，接收端同步于直射路径

(a) 跳频速率为多少时，系统呈现为无衰落？

(b) 考虑FFH系统，跳频间隔为$T_c=50\mu s$，码元间隔为$T_s=0.5\mu s$。问该系统呈现出的是无衰落、平衰落还是频率选择性衰落？

（c）考虑SFH系统，跳频间隔为T_c =50μs，码元间隔为T_s =0.5μs。问该系统呈现出的是无衰落、平衰落还是频率选择性衰落？

13.4 本题研究DSSS系统中，乘以扩频序列后的噪声的统计特性。设$n(t)$是自相关函数为$\rho_n(\tau)$的随机噪声过程，$s_c(t)$是均值为零、自相关函数为$\rho_c(\tau)$的随机扩频序列，$s_c(t)$它与$n(t)$相互独立。令 $n'(t)=n(t)s_c(t)$。

（a）求 $n'(t)$ 的自相关函数和功率谱密度。

（b）证明：如果$\rho_c(\tau)=\delta(\tau)$，则 $n'(t)$ 的均值为零、自相关函数为$\rho_n(\tau)$。就是说 $n'(t)$ 和$n(t)$有相同的统计特性，即乘以$s_c(t)$没有影响噪声的统计特性。

（c）如果$n(t)$是零均值的加性白高斯噪声，$s_c(t)$是最大线性码，其自相关函数由式（13-19）给出。试求$n'(t)$的自相关函数 $\rho_{n'}(\tau)$ 。当式（13-19）中的$N\to\infty$时，会发生什么情况？

13.5 证明：对于任意的实周期扩频码$s_c(t)$，它在一个周期上的自相关函数$\rho_c(\tau)$关于τ对称，且τ=0时最大。

13.6 假设$s_c(t)$的周期为T，证明$s_c(t)$的不同时延的相关值只与时延的差有关：

$$\frac{1}{T}\int_0^T s_c(t-\tau_0)s_c(t-\tau_1)\,\mathrm{d}t=\rho_c(\tau_1-\tau_0)$$

13.7 证明：对任意周期为T的扩频码$s_c(t)$，其自相关函数也是周期函数，且周期也为T。

13.8 证明：周期为$NT_c=T_s$的最大线性扩频码$s_c(t)$的功率谱密度为

$$P_{s_c}(f)=\sum_{m=-\infty}^{\infty}\frac{N+1}{N^2}\operatorname{sinc}^2\left(\frac{m}{N}\right)\delta\left(f-\frac{m}{T_s}\right)$$

并在N=100、T_s =1μs 的条件下，画出频率范围 $-10/T_s\leqslant f\leqslant 10/T_s$内的功率谱密度图。 444

13.9 证明m序列和随机二进制扩频序列一样具有平衡性、游程特性和移位特性。

13.10 用周期为T的最大线性码$s_c(t)$对未调载波$s(t)$进行扩频，然后通过冲激响应为$h(t)=\alpha_0\,\delta(t-\tau_0)+\alpha_1\,\delta(t-\tau_1)$的信道传输。接收信号$r(t)$送入图13-9中的同步环路。请写出图中积分器的输出$w(\tau)$。什么因素决定粗同步将锁定这两径的哪一径上？

13.11 假设有一个三分支的RAKE接收机，采用DPSK调制，各分支是独立的瑞利衰落，解扩前各支路上平均比特信噪比是10dB，扩频码是最大线性码，码长是$K=N=2^n-1=15$。假设第一分支理想同步，但第二分支偏了T_c /4，第三分支偏了T_c /3。在计算信噪比时，不考虑其他路径信号的干扰，并采用选择合并分集。试求P_b=10^{-6}所对应的中断率。

13.12 一扩频信号在多径信道中传输，信道包括一个直射径和一个延时路径，其中第二径相对于直射径的延时大于码片周期T_c。接收端采用两支路的RAKE接收机，分别对应这两个径。假设两个支路都理想同步时，每条支路解扩后的信号功率在6mW~10mW间均匀分布，解扩后的噪声功率为1mW。现假设仅第一支路是理想同步的，第二支路的时间偏移了T_c /2.366。假设扩频码的自相关函数如同N>>1的最大线性码。接收端采用最大比合并，合并器已知时钟偏移量。

（a）求合并输出的平均信噪比。

(b) 求合并输出信噪比的分布。

(c) 对于DPSK调制及误比特率10^{-4}，求中断率。

13.13 本题说明RAKE接收机的收益以及复杂度受限时多径分量的最佳选择问题。设多径信道的冲激响应为

$$h(t)=\alpha_0\delta(t)+\alpha_1\delta(t-\tau_1)+\alpha_2\delta(t-\tau_2)$$

其中α_i 为瑞利衰落系数。由于信道中存在阴影效应，系数α_i 的平均功率是变化的，$E[\alpha_0^2]$以对半的概率取值5或10，$E[\alpha_1^2]$也以相同的概率取值于0或20，$E[\alpha_2^2]$以概率0.75等于5，以概率0.25等于10（以上都是线性单位）。当接收机锁定在第i径时，若无其他径则发送功率和噪声功率恰好是输出信噪比等于α_i^2 。

445

(a) 假设采用最大线性码，比特间隔是T_b，扩频接收机与直射路径同步（延时为0，增益为α_0)。试求τ_1和$\tau_2(0\leqslant\tau_1\leqslant\tau_2\leqslant T_b)$的范围，使得在此范围内，这两径的分量将被衰减$-1/N$，$N$是每比特的码片数。

以下小题中假设扩频码的自相关函数是冲激函数，调制方式是DPSK，瞬时误比特率目标是$P_b=10^{-3}$。

(b) 若采用锁定在直射径的单支路扩频接收机，求中断率。

(c) 若采用三支路RAKE接收机，每条支路与多径中的一条同步，采用选择合并方式，求中断率。

(d) 假设接收机的复杂性受到限制，只能采用两支路的选择合并RAKE接收机。接收机应该锁定在哪两条径上，才能使中断率最小？并计算相应得中断率。

13.14 本题研究多径时延为随机时RAKE接收机的性能。一个直接序列扩频码的码片周期为T_c，自相关函数为

$$\rho_c(t)=\begin{cases}1 & -T_c/2<t<T_c/2\\ 0 & \text{其他情况}\end{cases}$$

用这个扩频码调制一个比特间隔为$T_b=10T_c$ 的DPSK信号，再通过多径信道传输。信道模型采用3.4节介绍的离散时间抽头模型，抽头间隔为T_c ，总的多径时延扩展为$T_m=5T_c$ 。这样的信道模型有5个多径位置，第i个位置至多有一个延时为$(i-0.5)T_c$的多径分量。不同位置上的多径分量相互独立的。在某一个位置上，观察到多径分量的概率是0.75。如果位置i上有多径分量，则解扩后第i径分量的幅度是瑞利分布，平均比特信噪比是$S_i=20/i$，$i=1, 2, \cdots, 5$（线性单位），即多径信号的平均功率传输距离的增大而减小。

接收端是一个五支路的选择合并RAKE接收机，每个支路与一个多径位置同步。假设目标误比特率为10^{-3}，请计算RAKE接收输出的中断率，并与这样一种情形进行比较：其中每个位置都有多径分量（即概率为1)，每个多径分量解扩后的平均比特信噪比是$S_i=20$（线性单位）。

13.15 由于宽带扩频信号有很好的多径分辨能力，直序扩频常用于测量信道，也称信道探测（channel sounding)。测量采用一个多支路的接收机，每个支路同步在不同的码片时延上，如图13-19所示。发送端发送一个未调制的扩频码$s_c(t)$，经过信道$h(t)$到达接收端。接

收端有$N+1$个支路，第i个支路的输出近似等于该时延上的信道增益，因此信道模型近似为 $\hat{h}(t)=\sum_{i=0}^{N}\beta_i\delta(t-iT_c)$ 。

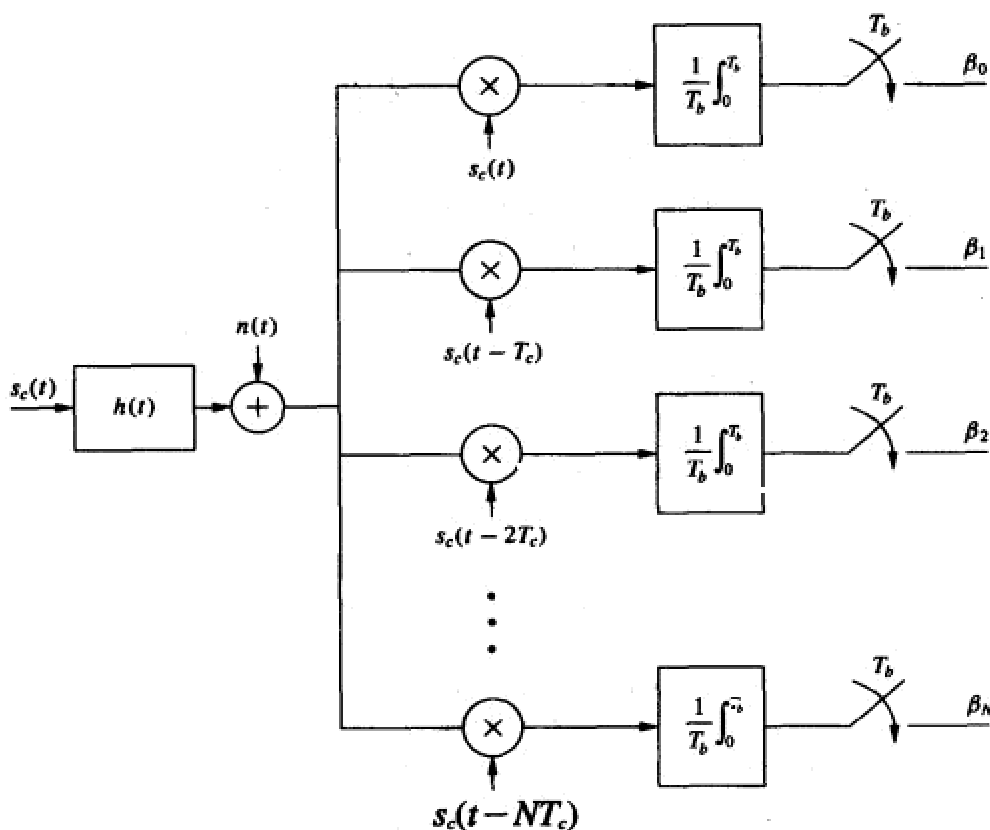


图 13-19

假设$s_c(t)$的自相关函数为

$$\rho_c(\tau)=\frac{1}{T_b}\int_0^{T_b}s_c(t)s_c(t-\tau)\,dt=\begin{cases}1-|\tau|/T_c & |\tau|<T_c\\ 0 & |\tau|\geqslant T_c\end{cases}$$

446

(a) 证明：如果 $\hat{h}(t)=\sum_{i=0}^{N}\alpha_i\delta(t-iT_c)$，那么没有噪声（$n(t)=0$）时，图13-19中各支路的输出就是$\beta_i=\alpha_i$。

(b) 仍不考虑噪声，若$h(t)=a\delta(t)+b\delta(t-T_c)+c\delta(t-3.5T_c)$，探测出 $\hat{h}(t)$是什么?

(c) 假设信道探测理想地探测出 $\hat{h}(t)=h(t)=\beta_0\delta(t)+\beta_1\delta(t-T_c)+\beta_2\delta(t-2T_c)$，其中$\beta_i$是独立的瑞利衰落随机变量。假设有一个三支路的RAKE接收机，三个支路理想同步于$h(t)$的三个多径分量，每支路解扩后的平均比特信噪比为10dB，调制方式为DPSK，目标误比特率为10^{-3}，RAKE接收机采用最大比合并，求P_{out}。若RAKE接收机采用选择合并，重新求P_{out}。

447

13.16 分别求出$n=8$时Gold码、小集合Kasami码、大集合Kasami码的自相关函数和互相关函数，并且求出大集合和小集合Kasami码的个数。

13.17　求N=4的Hardamard矩阵，并证明用该矩阵行形成的扩频码是正交的（即$\rho_{ij}(0)=0\forall i\neq j$）。若用户之间的时延差是$T_c/2$，求任意两用户之间互相关函数，即对任意的一对码，求

$$\rho_{ij}\left(\frac{T_c}{2}\right)=\frac{1}{T_s}\int_0^{T_s}s_{c_i}(t)s_{c_j}\left(t-\frac{T_c}{2}\right)\mathrm{d}t=\frac{1}{N}\sum_{n=1}^{N}s_{c_i}(nT_c)s_{c_j}\left(nT_c-\frac{T_c}{2}\right)$$

13.18　某异步DSSS多址接入信道（multiple access channel，MAC）有K=40个用户，扩频因子是$N=B_s/B$=100。假设该系统是干扰受限的，每个用户的信道都没有多径。采用标准高斯近似的随机码，用户信号经过BPSK调制。用户k受到了深衰落，使其到达接收端的接收功率比其他用户小6dB，试计算该用户的误比特率。各用户同步时误码率会不会改变？

13.19　对于式（13-54）给定的r_k，证明向量$\boldsymbol{r}=(r_1,\cdots,r_K)^T$可以写成式（13-56）的矩阵形式。式中$\boldsymbol{n}$的统计特性是什么？

13.20　对于K用户的同步MAC接收机，证明最大似然检测器选出的向量$\boldsymbol{b}$可以使式（13-57）中的代价函数最大。

13.21　本题说明，单用户CDMA系统可以用多个扩频码实现自适应调制和分集增益。采用速率为Rbit/s的BPSK调制时，K用户DS-CDMA系统中第k个用户的误比特率为

$$\mathrm{BER}_k=Q\left(\sqrt{\frac{2P_k(\gamma_k)\gamma_k}{\frac{1}{N}\sum_{i=1,i\neq k}^{K}P_i(\gamma_i)\gamma_i+1}}\right)\tag{13-61}$$

式中N是扩频因子（处理增益），γ_i为用户i的信道功率增益，$P_i(\gamma_i)$是用户i在其信道增益为γ_i时的发射功率。注意噪声功率已被归一化，接收端解调每一个扩频码时，把其他码视为噪声（匹配滤波检测）。系统中有单个用户，它最多可以用两个扩频码发送信号，每个码上的数据速率是Rbit/s，调制方式是BPSK。

（a）假设用户的信道衰落为γ，发送端把总的发射功率$P(\gamma)$在两个码上平分。用户有三种选择：不发送，发送一个BPSK调制的码，发送两个包含独立比特的BPSK调制的码。根据多用户情形下的误比特率公式（13-61）可以得到单用户多速率DS-CDMA系统中的误比特率。请据此说明用户只发一个码时的误码率是：

448

$$\mathrm{BER}=Q\left(\sqrt{2P(\gamma)\gamma}\right)\tag{13-62}$$

用户同时发两个码时的误比特率是：

$$\mathrm{BER}=Q\left(\sqrt{\frac{P(\gamma)\gamma}{P(\gamma)\gamma/2N+1}}\right)\tag{13-63}$$

这两种情况下可达到的数据速率是多少？

（b）假设收发都理想已知信道，γ的分布是$p(\gamma)$。我们来设计一种自适应速率自适应功率的方案。不采用纠错编码，用户要求所有发送比特的误比特率都必须低于门限P_b^0。假设

PDG

平均发射功率归一化：$\int_0^\infty P(\gamma)p(\gamma)\mathrm{d}\gamma=1$。可能的发送速率是有限的，因此与窄带自适应调制类似，最佳的自适应方案是：当γ低于截止门限γ_0时，不发数据；当$\gamma_0\leqslant\gamma<\gamma_1$时只发一个码；当$\gamma>\gamma_1$时发两个码。对于这种自适应速率方案以及给定的门限$\gamma_0$和$\gamma_1$，试找出能严格满足目标误比特率要求$P_b^0$的自适应功率分配方案。

（c）对于（b）中得到的自适应速率和自适应功率方案下，用拉格朗日法求出γ_0和γ_1作为拉格朗日系数λ的函数。

13.22 某公司想设计一个能支持话音和高速数据的下一代蜂窝系统。FCC已决定给该系统分配100MHz的带宽，不管产业界达成的具体标准是什么。你负责设计这个系统并向标准化组织推出设计方案。你必须尽可能详细地说明你的设计，着重要说明怎样克服衰落和ISI的影响以及容纳话音和数据业务的能力。还要论证为什么你的设计方案要比其他设计更好。

参考文献

[1] R. Pickholtz, D. Schilling, and L. Milstein, "Theory of spread-spectrum communications – A tutorial," *IEEE Trans. Commun.*, pp. 855–84, May 1982.

[2] R. L. Peterson, R. E. Ziemer, and D. E. Borth, *Introduction to Spread Spectrum Communications*, Prentice-Hall, Englewood Cliffs, NJ, 1995.

[3] V. V. Veeravalli and A. Mantravadi, "The coding–spreading trade-off in CDMA systems," *IEEE J. Sel. Areas Commun.*, pp. 396–408, February 2002.

[4] S. W. Golomb, *Shift Register Sequences*, Holden-Day, San Francisco, 1967.

[5] A. J. Viterbi, *CDMA Principles of Spread Spectrum Communications*, Addison-Wesley, Reading, MA, 1995.

[6] G. L. Stuber, *Principles of Mobile Communications*, 2nd ed., Kluwer, Dordrecht, 2001.

[7] O. C. Mauss, F Classen, and H. Meyr, "Carrier frequency recovery for a fully digital direct-sequence spread-spectrum receiver: A comparison," *Proc. IEEE Veh. Tech. Conf.*, pp. 392–5, May 1993.

[8] R. C. Dixon, *Spread Spectrum Systems with Commercial Applications*, 3rd ed., Wiley, New York, 1994.

[9] M. K. Simon, J. K. Omura, R. A. Scholtz, and B. K. Levitt, *Spread Spectrum Communications Handbook*, McGraw-Hill, New York, 1994.

[10] B. Sklar, *Digital Communications – Fundamentals and Applications*, Prentice-Hall, Englewood Cliffs, NJ, 1988.

449

[11] G. L. Turin, "Introduction to spread spectrum antimultipath techniques and their application to urban digital radio," *Proc. IEEE*, pp. 328–53, March 1980.

[12] M. K. Simon and M.-S. Alouini, *Digital Communication over Fading Channels: A Unified Approach to Performance Analysis*, Wiley, New York, 2000.

[13] R. Gold, "Optimum binary sequences for spread-spectrum multiplexing," *IEEE Trans. Inform. Theory*, pp. 619–21, October 1967.

[14] E. H. Dinan and B. Jabbari, "Spreading codes for direct sequence CDMA and wideband CDMA cellular networks," *IEEE Commun. Mag.*, pp. 48–54, September 1998.

[15] N. B. Mehta, L. J. Greenstein, T. M. Willis, and Z. Kostic, "Analysis and results for the orthogonality factor in WCDMA downlinks," *IEEE Trans. Wireless Commun.*, pp. 1138–49, November 2003.

[16] S. Verdú, "Demodulation in the presence of multiuser interference: Progress and misconcep-

tions," in D. Docampo, A. Figueiras, and F. Perez-Gonzalez, Eds., *Intelligent Methods in Signal Processing and Communications*, pp. 15–46, Birkhäuser, Boston, 1997.
[17] S. Verdú, *Multiuser Detection*, Cambridge University Press, 1998.
[18] M. Pursley, "Performance evaluation for phase-coded spread-spectrum multiple-access communication – Part I: System analysis," *IEEE Trans. Commun.*, pp. 795–9, August 1977.
[19] R. Pickholtz, L. Milstein, and D. Schilling, "Spread spectrum for mobile communications," *IEEE Trans. Veh. Tech.*, pp. 313–22, May 1991.
[20] K. S. Gilhousen, I. M. Jacobs, R. Padovani, A. J. Viterbi, L. A. Weaver, Jr., and C. E. Wheatley III, "On the capacity of a cellular CDMA system," *IEEE Trans. Veh. Tech.*, pp. 303–12, May 1991.
[21] H. Xiang, "Binary code-division multiple-access systems operating in multipath fading, noisy channels," *IEEE Trans. Commun.*, pp. 775–84, August 1985.
[22] S. Verdú, "Optimum multiuser signal detection," Ph.D. thesis, University of Illinois, Urbana-Champaign, August 1984.
[23] S. Verdú, "Minimum probability of error for asynchronous Gaussian multiple-access channels," *IEEE Trans. Inform. Theory*, pp. 85–96, January 1986.
[24] A. Duel-Hallen, J. Holtzman, and Z. Zvonar, "Multiuser detection for CDMA systems," *IEEE Pers. Commun. Mag.*, pp. 46–58, April 1995.
[25] R. Lupas and S. Verdú, "Linear multiuser detectors for synchronous code-division multiple-access channels," *IEEE Trans. Inform. Theory*, pp. 123–36, January 1989.
[26] M. L. Honig and H. V. Poor, "Adaptive interference suppression," in *Wireless Communications: Signal Processing Perspectives*, ch. 2, Prentice-Hall, Englewood Cliffs, NJ, 1998.
[27] S. Verdú and S. Shamai (Shitz), "Spectral efficiency of CDMA with random spreading," *IEEE Trans. Inform. Theory*, pp. 622–40, March 1999.
[28] M. K. Varanasi and B. Aazhang, "Near-optimum detection in synchronous code division multiple-access communications," *IEEE Trans. Commun.*, pp. 725–36, May 1991.
[29] M. K. Varanasi and B. Aazhang, "Multistage detection in asynchronous code division multiple-access communications," *IEEE Trans. Commun.*, pp. 509–19, April 1990.
[30] A. Duel-Hallen, "Decorrelating decision-feedback multiuser detector for synchronous CDMA," *IEEE Trans. Commun.*, pp. 285–90, February 1993.
[31] A. Duel-Hallen, "A family of multiuser decision-feedback detectors for asynchronous code-division multiple-access channels," *IEEE Trans. Commun.*, pp. 421–34, February–April 1995.
[32] J. G. Andrews, "Interference cancellation for cellular systems: A contemporary overview," *IEEE Wireless Commun. Mag.*, pp. 19–29, April 2005.
[33] D. Divsalar, M. K. Simon, and D. Raphaeli, "Improved parallel interference cancellation for CDMA," *IEEE Trans. Commun.*, pp. 258–68, February 1998.
[34] Y. C. Yoon, R. Kohno, and H. Imai, "A spread-spectrum multiaccess system with cochannel interference cancellation for multipath fading channels," *IEEE J. Sel. Areas Commun.*, pp. 1067–75, September 1993.
[35] P. Patel and J. Holtzman, "Analysis of a simple successive interference cancellation scheme in a DS/CDMA system," *IEEE J. Sel. Areas Commun.*, pp. 796–807, June 1994.
450
[36] A. J. Viterbi, "Very low rate convolutional codes for maximum theoretical performance of spread-spectrum multiple-access channels," *IEEE J. Sel. Areas Commun.*, pp. 641–9, May 1990.
[37] J. G. Andrews and T. H. Meng, "Optimum power control for successive interference cancellation with imperfect channel estimation," *IEEE Trans. Wireless Commun.*, pp. 375–83, March 2003.
[38] X. Wang and H. V. Poor, "Blind equalization and multiuser detection in dispersive CDMA channels," *IEEE Trans. Commun.*, pp. 91–103, January 1998.
[39] U. Mitra, "Comparison of maximum-likelihood-based detection for two multirate access schemes for CDMA signals," *IEEE Trans. Commun.*, pp. 64–77, January 1999.
[40] X. Wang and H. V. Poor, *Wireless Communication Systems: Advanced Techniques for Signal Re-*

ception, Prentice-Hall, Englewood Cliffs, NJ, 2004.

[41] R. Kohno, R. Meidan, and L. B. Milstein, "Spread spectrum access methods for wireless communications," *IEEE Commun. Mag.*, pp. 58–67, January 1995.

[42] H. El Gamal and E. Geraniotis, "Comparing the capacities of FH/SSMA and DS/CDMA networks," *Proc. Internat. Sympos. Pers., Indoor, Mobile Radio Commun.*, pp. 769–73, September 1998.

451



第14章　多用户系统

在多用户系统中，系统资源必须要分配给不同的用户。本章讨论将系统资源分配给多个用户的技术以及多用户系统的容量极限。从5.1.2节中我们知道，一个带宽为B、周期为T的信号的空间维数是$2BT$。为了支持多用户通信，必须要将系统的信号空间维分配给不同的用户[1]，这样的分配称为*多址接入*[2]（multiple access）。不同的多址接入技术在不同多用户信道下的性能也不同，我们将主要考虑两种基本的多用户信道：上行信道和下行信道。将信号空间维分配给不同用户的方式有无限多种，因此多用户信道的容量不是用一个数来衡量，而是用*速率域*（rate region）。速率域是各用户的速率构成的向量的集合，如果各用户按速率域中的速率向量进行传输，所有用户的错误概率都可以任意小。本章将讨论上行和下行多用户信道的容量域。同时也要讨论只将信号维分配给激活用户的随机接入技术，以及能保证用户达到可接受性能所需SINR的功率控制技术。本章还将讨论多用户分集，它能借助不同用户信道的时变性。本章最后讨论多用户系统采用多天线时的信号技术及性能增益。

452

14.1　多用户信道：上行与下行

多用户信道指任何由多个用户所共享的信道，它有两种不同的类型：上行信道和下行信道，如图14-1所示。下行信道是一个发送端对多个接收端，也称为广播信道或前向信道。到达所有接收机的信号都来自同一个下行的发射机，所以发射信号是发给所有K个用户的信号之和：$s(t)=\sum_{k=1}^{K}s_k(t)$，信号的总功率为$P$，总带宽为$B$。总的信号维和功率在多个用户之间分配。所有用户信号都从一处发射，所以下行容易实现用户间的同步，不过多径有可能破坏这种同步。下行的另一个重要的特点是信号和干扰都经历了相同的信道失真。例如用户k的信号$s_k(t)$和其他干扰信号$s_j(t), j\neq k$都经过用户k的信道$h_k(t)$到达用户k的接收端。这是上行信道和下行信道的本质差别：上行信道中，来自不同用户的信号所经过的是不相同的信道。广播电台、电视广播、卫星到地面站、蜂窝系统中基站到移动台都是下行信道。

上行信道也称为多址信道[3]或反向信道，它是多个发送端对一个接收端，每一个发射信号都在系统带宽B内。与下行信道不同，每个上行用户以单独的功率约束P_k发送信号$s_k(t)$。此外，由于发送的信号来自于不同的发射机，如果要求所有信号同步，就需要各发射机能协同工作。图14-1还示出了不同上行信号所经历的不同信道，因此即使发送功率相同，如果各用户的信道增益不同的话，接收功率也不同。无线局域网卡到接入点、地面站到卫星、蜂窝系统中移动台到基站都是上行信道。

453

1. 通过多址或随机接入对信号空间维的分配在OSI模型的MAC层完成。
2. 正如在14.5节中所讨论的，分配给不同用户的信号空间维不一定是正交的。
3. 注意多址技术对上行信道和下行信道都是适用的。

PDG

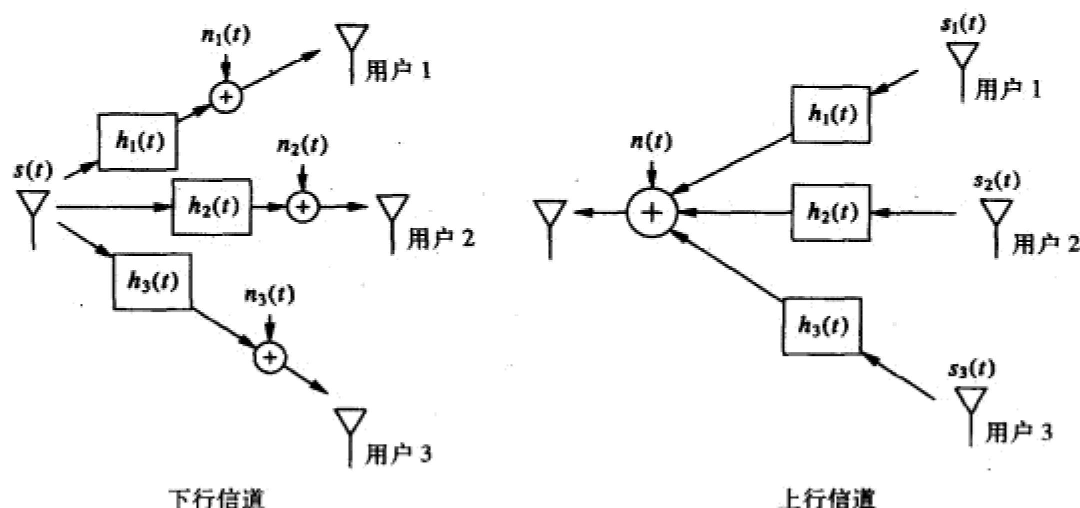


图14-1 下行信道和上行信道

大多数通信系统是双向的，既有上行信道也有下行信道。通过下行信道向用户发送，通过上行信道接收用户信号的无线收发信机称为接入点或基站。由于干扰的原因，接收和发送一般不能同时在同一频带上，需要通过正交信号维来分离上行和下行信道，称为双工（duplexing），一般采用时间维分离或频率维分离。时分双工（Time-division duplexing，TDD）在正交的时隙上进行收发；频分双工（frequency-division duplexing，FDD）在不同的频带上进行收发。TDD的优点是双向的信道增益一般是对称的，在一个方向上的信道测量可以作为另一个方向的信道估计。FDD在频率选择衰落不一定是这种情况，如果上下行频率间隔大于多径信道的相干带宽，那么两个信道的衰落是独立的。

14.2 多址接入

带宽是宝贵的资源，因此上下行信道设计的一个关键就是如何高效地分配信号维。给每个用户分配专用信道方式常称为多址接入[1]。连续传输且时延受限的业务，如语音或视频，需要专用信道以保证传输质量。得到专用信道的方法是对信号空间进行时分、频分、码分或者它们的结合。对于那些突发传输的用户，通常采用某种随机形式的信道分配将信号维分配给用户，但这样并不能保证信道的接入，这种方式称为随机多址或随机接入（random access），14.3节中将进一步介绍。一个系统应该采用多址接入还是随机接入，应该采用哪种具体的接入形式，这些问题一般取决于系统的具体应用、用户的业务特性、系统的性能要求、信道的特性以及同频带上其他干扰系统的特性等等因素。

454

多址接入技术将信号维划分为不同的信道后分配给用户。一般是按时间轴、频率轴或码轴将信号空间的维分割为正交或非正交的用户信道。时分多址（time-division multiple access，TDMA）和频分多址（frequency-division multiple access，FDMA）是

1. 上行信道也称为多址信道，但上行下行都需要多址技术。

正交信道，码分多址（code-division multiple access，CDMA）根据扩频码设计的不同，可以是正交的或非正交的。用天线阵列或其他式产生的有向天线也能使信号空间增加了一个角度维，利用这个维划分信道就是空分多址（space-division multiple access，SDMA）。不同多址方式的性能与它们各自的特性以及是用于上行还是下行有关。TDMA、FDMA和正交CDMA在正交划分信号维的意义下是等价的，它们所能产生出的正交信道数相同。对于维数为$2BT$的信号空间，无论采用何种信道划分方法，总可以分为N个包含$2BT/N$维的信道。所有正交划分的多址方式在AWGN信道中都有相同的信道容量，14.5节和14.6节将对此进行讨论。但信道中可能存在的平衰落或频率选择性衰落等不理想因素对不同多址方式的影响是不同的，这将使不同的多址方式有不同的信道容量和性能。

14.2.1 频分多址（FDMA）

FDMA沿频率轴将信号维划分成不重叠的信道，如图14-2所示，每个用户分配一个不同的频率信道。信道之间一般设有保护带，以补偿滤波特性的不理想、邻信道干扰和多普勒扩展。如果信道非常窄，那么即使总带宽很宽，每个信道也不会经历频率选择性衰落。FDMA中，信号在时间上连续发送，这有可能使开销问题复杂化，比如信道估计的开销必须要同时在整个带宽内传输。FDMA也要求无线收发信机的频率可调谐，能根据需要调到不同信道的载频上。此外，在FDMA系统中，给一个用户同时分配多个信道有一定的难度，这要求收信机能进行多频点接收。尽管如此，模拟通信一般都采用FDMA作为其多址技术，例如AMPS和TACS模拟蜂窝系统（见文献[2]的11.1节）。OFDM系统中的多址技术OFDMA也是一种FDMA，它将不同的子载波分配给不同的用户。

455

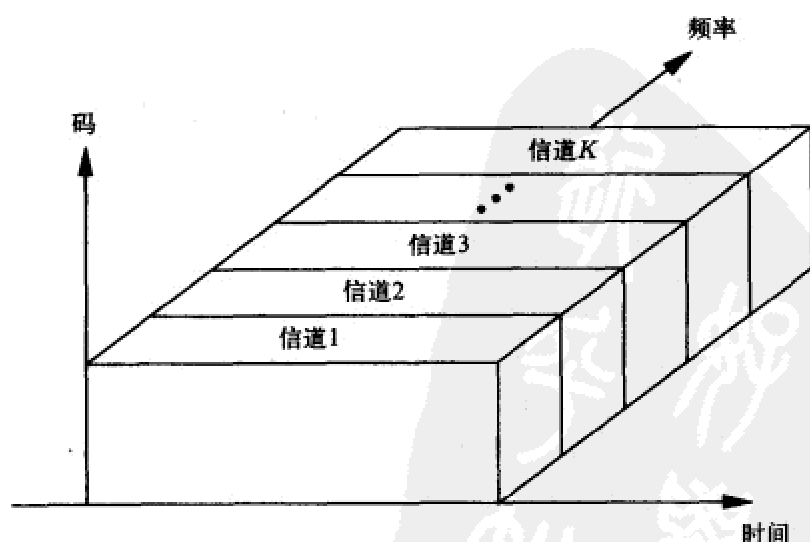


图14-2 频分多址

例14.1：第一代模拟系统分配的总带宽是上下行各25MHz。这25MHz的带宽平分给两个运营商，每个运营商的带宽是上下行各12.5MHz。分配给每个用户的带宽是B_c= 30kHz以支持模拟话音业务，它的上下行都采用带宽为24kHz的FM调制，两边各留3kHz的保护带。为了减小邻系统之间的干扰，整个上下行频段的两边各留B_g= 10kHz的保护带。试计算这25MHz带宽总共能容纳的模拟话音用户数。再考虑一个效率更高的数字系统，它采用了高阶调制，使得每路数字话音只需要10kHz的带宽；它还采用了更严格的滤波，使得频带两侧的保护带只需要5kHz。对于同样的上下行25MHz带宽，数字系统最多能容纳多少话音用户？

解：无论是上行还是下行，考虑到保护带后，N个用户需要的总的带宽为NB_c+2B_g，因此对于上下行信道带宽B=25MHz，系统所能支持的用户数为

$$N=\frac{B-2B_g}{B_c}=\frac{25\times10^6-2\times10\times10^3}{30\times10^3}=832$$

每运营商为416个用户，即每小区总共可支持832个用户。对于数字系统，每小区所能支持的用户数为

$$N=\frac{B-2B_g}{B_c}=\frac{25\times10^6-2\times5\times10^3}{10\times10^3}=2499$$

是模拟系统的三倍。容量提高的主要原因是采用了高阶调制，它使数字话音信号需要的带宽仅是模拟话音信号的三分之一。

14.2.2 时分多址（TDMA）

TDMA系统沿时间轴分割信号维，分配给用户的信道是互不重叠并周期重复的时隙，如图14-3所示。每个TDMA信道都要占用整个带宽，总带宽一般很宽，因此需要采用一些对抗ISI的技术。时隙循环重复使用户的信号不是连续发送，需要采用可以缓存的数字传输方式。不连续发送简化了开销，例如信道估计可以在其他用户发送的时候进行。TDMA还有一个优点是，只要简单地分配多个时隙，就能使一个用户拥有多个信道。

456

TDMA的主要问题是同步，至少对上行如此。下行的所有信号都是同一发送端发出的，到达任一接收机时经历了同样的信道，因此平衰落时，只要用户在正交时隙中发送，到接收端还是正交的。上行的接收信号来自各个发送端，各用户的发送时机以及发送信号经历的信道都不相同，要想使接收到的信号保持时间正交，就必须要使所有用户同步发送。一般靠基站或接入点的协调来实现同步，往往会有较大的开销。无论是上行还是下行，多径也有可能破坏时分的正交性，如果多径时延和时隙长度相比不可忽略的话。为此，TDMA一般在时隙之间设有保护间隔以减小同步误差和多径的影响。TDMA的另一个问题是，发送时隙是循环重复的，循环到下一次时，信道特性可能已经改变。这对于均衡等需要信道估计的功能来说，就必须要在每个循环周期内重新进行信道估计，而连续发送时只需要进行跟踪。时分多址技术已用于GSM、PCD和IS-136数字蜂窝系统，见文献[2]的第11章。

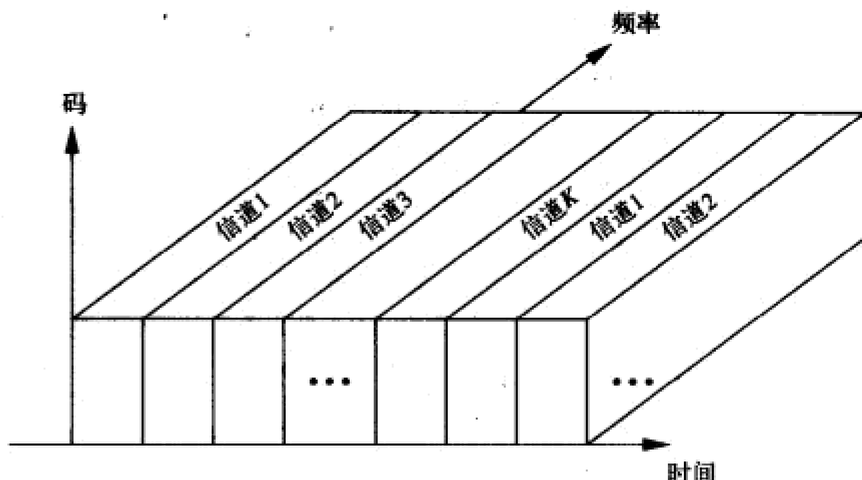


图14-3 时分多址接入

例14.2：GSM的原始设计和AMPS一样，上下行都有25MHz带宽。这25MHz带宽分为125个200kHz宽的信道，每个信道又分为8个用户时隙，这8个时隙加上前导和后置比特构成一帧，帧按时间循环重复。试计算GSM能支持的用户数和每个用户的信道带宽。如果信道的均方根时延扩展为10μs，请问该系统是否需要采取抵抗ISI的措施？

解：每个信道有8个用户，共有125个信道，可容纳的用户数是125 × 8=1000个，每个TDMA信道的带宽为 $\frac{25\times10^6}{125}=200\text{kHz}$ 。10μs的时延扩展对应的信道相干带宽为B_c=100kHz，小于信道带宽200kHz，因此需要考虑减小ISI的问题。GSM规范包括了均衡器，具体均衡器的类型则由设计者决定。

457

14.2.3 码分多址（CDMA）

CDMA用正交或非正交的扩频码来调制信息信号，不同用户的扩频信号占用相同的时间相同的频带，如图14-4所示。接收端利用扩频码的结构分离出不同的用户。CDMA常见的形式就是13.4节和13.5节所介绍的直序和跳频的多用户扩频。

CDMA的下行一般采用正交扩频码，如Walsh-Hadamard码，虽然多径会破坏其正交性。上行一般采用非正交的扩频码，主要是因为同步比较困难，并且在多径信道中维持上行正交也很复杂[3]。上行信道采用非正交码的好处是完全靠码区分用户，基本不需要再在时间和频率上进行协调。此外，TDMA和FDMA的信道都对信号维进行正交划分，可分出的信道数有硬限制，使用正交码的CDMA也一样，而使用在非正交码的CDMA中信道数没有硬限制。不过，非正交码会带来用户间干扰，用户越多干扰也越大，这个干扰会使全体用户的性能恶化。非正交CDMA也需要上行功率控制以避免远近效应。之所以会出现远近效应，是因为各用户的上行信道增益不同时，到达接收端的功率也不同。假设有两个用户，一个离基站或接入点很近，另一个很远，两个用户都以相同的功率发送，那么近处用户产生的干扰将会淹没远处用户的信号。引入功率控制可以使各用户到

达基站的接收功率大致相同。这种信道反转式功控使各个干扰源在接收端的贡献相同，从而消除了远近效应。非正交CDMA系统也可以采用多用户检测技术来减小用户间干扰。即使在理想功控下，多用户检测技术也能使系统的性能得到可观的提高，如果把功控和多用户检测进行联合优化，性能还会更好[4]。在14.5节和14.6节中我们将看到，有一种结合了多用户检测的CDMA可以达到上行信道和下行信道的香农极限，虽然这两种信道下达到极限的方法有很大的区别。CDMA系统中，只需给同一用户分配多个不同的码字，就可以实现一个用户有多个信道。IS-95数字蜂窝标准采用了CDMA，其下行信道采用正交码，上行信道采用正交码和非正交码的结合，见文献[2]的11.4节。W-CDMA和CDMA2000标准也采用了CDMA的多址方式，见文献[5]的10.5节。

458

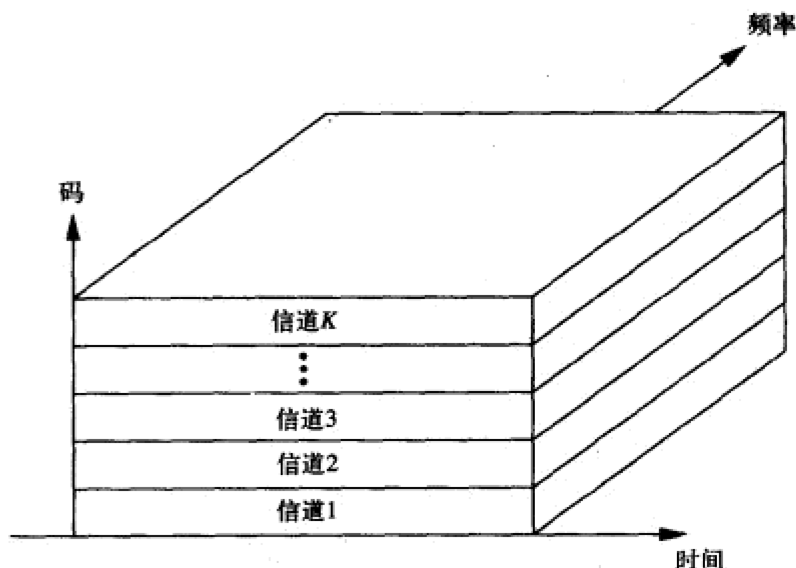


图14-4 码分多址

例14.3：根据式（13-47）的标准高斯近似，非正交CDMA上行信干扰（SIR）为

$$\mathrm{SIR}=\frac{3G}{K-1}$$

其中K是用户数，$G\approx 128$是扩频因子。IS-95上行信道的带宽是1.25MHz，因此扩频前信号的带宽为$B_s\approx\frac{1.25\times10^6}{128}=9.765\text{kHz}$。忽略噪声，若要求SIR为10dB，CDMA的上行能容纳多少用户？同样带宽的FDMA能容纳多少用户？

解：为了得到可支持的用户数，我们将SIR取倒数，得到

$$K\leqslant\frac{3G}{\mathrm{SIR}}+1=\frac{384}{10}+1=39.4$$

K必须是整数，故此系统可支持39个用户。对于FDMA，可得到

$$K=\frac{1.25\times10^6}{9.765\times10^3}=128$$

故此，1.25MHz带宽可支持128个9.765kHz的信道。计算结果表明，如果CDMA扩频码的互相关满足标准高斯近似，那么它的容量大约是FDMA的三分之一（还可算出，其他互相关假设下FDMA的容量更高）。但实际的IS-95上下行都能支持64个用户，它利用了能适应干扰情况和信道质量的可变压缩率话音编码，并利用了信道干扰并不总存在的特点（话音激活因子）。即使是这样，单小区的FDMA容量也比单小区的CDMA大。但蜂窝系统是要进行信道复用的，在这方面，CDMA效率要比FDMA高得多，15.2节将讨论这一点。

459

14.2.4　空分多址（SDMA）

空分多址将用户的方向看作信号空间的另外一个可以划分的维，一般是用有向天线来实现空间的信道划分，如图14-5所示。仅当两个用户的角度差大于天线的分辨角时，才能实现正交的空间信道。如果天线的方向性是用天线阵列来实现的，那么精确分辨空间角度需要很大的阵列，这在基站和接入点是不太现实的，对体积很小的用户终端更不可能。常见的SDMA形式是10.8节所介绍的扇区化天线阵列，它将360°的角度分成N个扇区，扇区内有很高的天线增益，扇区间干扰很小。每个扇区中可以继续用TDMA或FDMA等多址方式来区分用户。对于移动的用户，SDMA应能跟踪用户的角度变化，对于扇区式SDMA就是在用户移出一个扇区时，应能切换到另一个扇区。

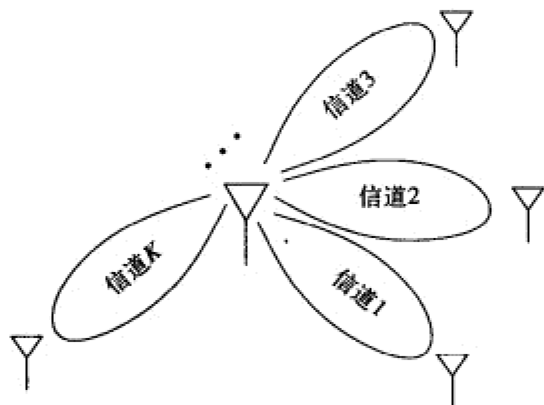


图14-5　空分多址

14.2.5　混合多址

许多系统采用不同多址方式的组合来分配信号维。OFDMA可以和子载波跳频结合以提高频率分集增益[6]。DSSS也可以结合FDMA使用以将系统带宽分成多个子带。这种方法将不同的用户分配到不同的子带上，再在子带内扩频。比起在整个系统带宽内的扩频，子带的处理增益较小，抗干扰和抗ISI能力也较低。但它的好处是子带之间不必是连续频谱，同时可以根据需要灵活地把信号扩频到不同带宽的子带上。另一种混合多址技术是将DS-CDMA和FH-CDMA结合，扩频信号的载频在一定带宽内跳变。干扰用户也在跳频，因此这种方法能减小远近效应。TDMA也可以和FH结合，这样，用户只会在周期性跳到的时候遇到深衰落或强干扰的信道，进而通过纠错编码就能消除其影响。GSM标准就采用了这种方法来降低邻小区的强干扰。

460

对于现有的和未来的无线通信系中的不同多址技术的性能比较，人们已经进行了许多的研究、讨论乃至争论（见文献[6~12]）。虽然可以对简单的系统在简单的信道模型下进行分析，并得到一般性的结论，但对于运行在许多典型条件下的复杂多用户系统，

很难说哪一种多址技术最好。而且，为了进行比较分析和仿真研究，不得不采用一些简化的假设条件，这样的假设条件有可能对某种特定的技术有利。与大多数工程设计问题一样，多址技术的选择取决于系统要求、系统特性、成本和复杂性的限制等因素。

14.3 随机接入

多址接入技术主要应用在语音、视频这类需要连续发送的业务中，对这些业务来说，专用信道可以实现好的性能。然而，大部分数据业务中，数据在随机的时间出现，并不需要连续传输。如果给这样的业务分配一个专用的信道，显然是非常低效的。此外，大多数系统中总的用户数（激活用户和闲置用户）比系统能同时容纳的用户数多很多，因此任何时间的信道分配只能是谁需要就分给谁。这样的系统采用随机接入策略就能高效地将信道分配给激活用户。

所有随机接入技术都是基于分组数据或称*分组无线电*（packet radio）的。它将用户数据组织成N比特的数据分组，其中可能包括检错、纠错和控制比特，然后通过信道传输。对于速率为Rbit/s的信道，一个分组的传输时间为$\tau = N/R$，R是占用信道全部带宽时的速率，不包含多用户发送时必须要有的用于区分用户的信令。如果不同用户所发送的分组在时间上重叠，就发生了碰撞，此时这两个分组都可能被译错。信道噪声和信道失真也可能造成译错。分组被译错的概率称为*分组错误率*（packet error rate）。随机接入的理论分析中，一般假设用户全体以泊松过程产生分组，单位时间内产生的分组数是λ，它是任意时间段$[0, t]$内的平均分组数除以t，任意时间段$[0, t]$内产生的平均比特个数除以t就是λN。对于泊松过程，$[0, t]$内到达的分组数$X(t)$等于整数k的概率为

$$p(X(t)=k)=\frac{(\lambda t)^k}{k!}e^{-\lambda t} \tag{14-1}$$

泊松过程是无记忆的，任一时间段内到达的分组数与其他时间段内到达的分组数无关。需要注意的是，泊松分布不一定对所有用户业务都是最好的模型，特别在互联网中，高突发的数据往往造成分组到达呈现出相关性[13]。

461

给定速率为λ的泊松到达以及分组传输时间τ，信道的业务负载定义为$L = \lambda\tau$。如果信道数据速率为每秒R_p个分组，则$\tau = 1/R_p = N/R$，其中R是以每秒比特为单位的信道速率。注意L是无单位的，它是分组到达的速率除以信道全速传输时能够传输的速率。$L>1$表示给定时间内平均到达的分组数要多于在同样时间内能够发出的分组数，假设碰撞必然引起错误，则$L>1$时系统是不稳定的。如果发生错误时接收端能通知发送端重传错误分组，那么分组到达率λ以及相应的负载的计算应包括新到的分组和需要重传的分组，此时的L称为*总提交负载*（total offered load）。

随机接入的性能通常用*吞吐量*（throughput）T反映。T没有单位，定义为给定时间内平均的成功发送的分组速率除以信道的分组传输速率R_p，也等于提交负载乘以成功接收分组的概率，即$T=Lp$（成功接收分组），这个概率与所用的随机接入协议以及信道特性有关，即使无碰撞，信道也会引起分组出错。因此$T\leqslant L$。由于$L>1$时系统不稳定，

所以稳定系统满足$T \leqslant L \leqslant 1$。可以看出吞吐量和信道的数据速率$R$无关，这是因为负载和相应的吞吐量都已被信道的数据速率R归一化。这就使随机接入协议的分析结果对任意链路设计和信道容量是通用的。对于链路数据速率为Rbit/s的分组无线电系统，由于T是以速率R发送时分组或比特成功的比例，因此有效数据速率是RT。随机接入的设计目标是让T尽可能大，以充分利用链路速率R。注意，有时候分组在时间上重叠不一定表示发生碰撞。比如短暂重叠，到达分组有不同的信道增益，有纠错编码等，此时一个或多个分组有可能被正确接收，这种情况称为*捕获效应*（capture effect），见文献[14]的4.3节。

Abramson提出的ALOHA协议开创了随机接入技术[15]。它把数据进行分组，一旦用户有数据需要发送，就发送这个分组。用户之间的碰撞使ALOHA的效率很低，吞吐量很小。采用划分时隙并使用户同步的方法可以成倍提高吞吐量，但即便如此，碰撞也使系统的吞吐量相对较低。为了避免碰撞从而增加吞吐量，人们对ALOHA协议提出了一些改进，这些方法包括载波侦听、碰撞检测、碰撞避免等。对于长突发的分组，可以通过调度来避免碰撞，但这样做需要额外的开销。本节将介绍各种随机接入技术、它们的性能及设计问题。

462

14.3.1　纯ALOHA

在纯ALOHA（非时隙ALOHA）中，用户产生出分组后立即发送。若不考虑捕获效应，则可认为重叠的分组一定会出错，必须要重发。假设没有发生碰撞的分组一定能正确接收（即信道中没有噪声和失真），则吞吐量等于提交负载乘以不发生碰撞的概率，即$T=Lp$（无碰撞）。假设一个用户在时间$[0, \tau]$内发送一个持续时间为τ的分组，那么当其他用户在$[-\tau, \tau]$内也发送一个持续时间同样为τ的分组时就会发生碰撞。不发生碰撞的概率就是时间$[-\tau, \tau]$内没有分组到达的概率，即式（14-1）中$t=2\tau$时的概率：

$$p(X(t)=0)=\mathrm{e}^{-2\lambda\tau}=\mathrm{e}^{-2L} \tag{14-2}$$

相应的吞吐量为

$$T=L\mathrm{e}^{-2L} \tag{14-3}$$

如图14-6所示，随着负载逐步增加，吞吐量也逐步增加，在$L=0.5$时达到最大值0.18，随后开始下降。也就是说，最大的数据速率仅为单用户连续发送时的18%。出现这个极值的原因在于，L很小时，信道大量空闲，吞吐量很小。随着L不断增大，信道的使用率增加，但碰撞也开始增加。$L=0.5$时达到了一种最佳平衡：用户产生的数据分组足够使信道有合理的使用率，同时分组产生的速率又不至于使碰撞太过频繁。$L>0.5$后，碰撞越来越频繁，吞吐量逐步下降。当L很大时，大部分的分组都会发生碰撞，吞吐量趋近于零。

纯ALOHA效率不高的部分原因是用户可以在任意时刻发送分组，两个或多个分组发生局部重叠就能破坏所有分组的接收。让用户同步，使其发射时间对齐就能避免这种局部重叠，这就是时隙ALOHA的原理。

463

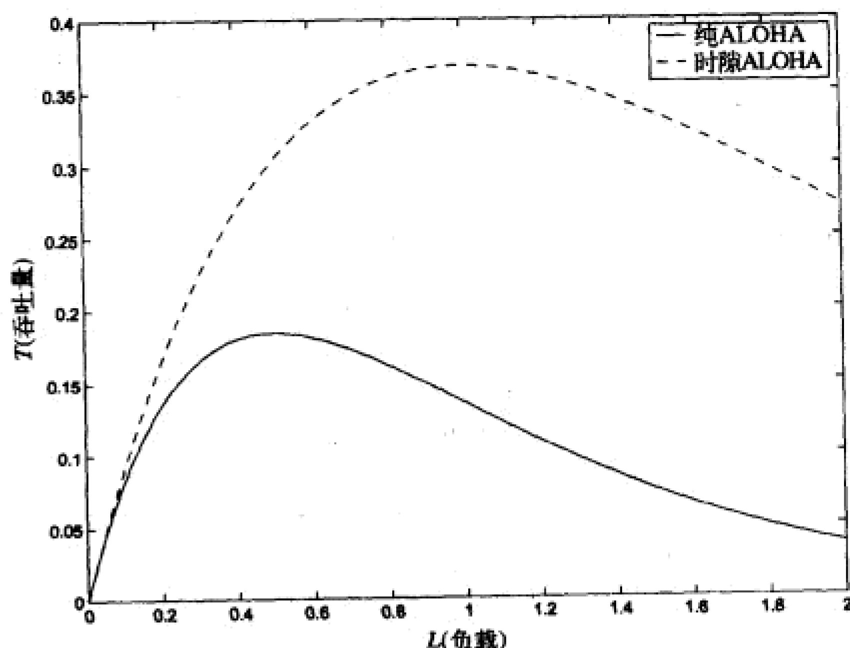


图14-6 纯ALOHA和时隙ALOHA的吞吐量

14.3.2 时隙ALOHA

时隙ALOHA把时间划分成持续时间为τ的时隙，用户准备好分组后，必须要等到下一个时隙的起始点才能开始发送分组，因此发送的数据分组不会发生局部重叠，这样就可以提高吞吐量。若一个分组在时间$[0, \tau]$内发送，这段时间内没有其他分组发送时，就能成功接收，其概率可将$t = \tau$代入式（14-1）得到为$p(X(t) = 0) = e^{-L}$，相应的吞吐量为：

$$T = Le^{-L} \tag{14-4}$$

如图14-6所示，随着负载逐步增加，吞吐量也逐步增加，在L=1时达到最大值T=0.37，随后开始下降。时隙ALOHA的最大吞吐量是纯ALOHA的两倍，并且在更高的负载下达到这个最大值。虽然这已表明时隙ALOHA的吞吐量已有显著提高，但它的有效数据速率仍然低于信道数据速率的40%。这对于紧张的无线带宽来说，仍然是一种严重的浪费，需要更好的技术来提高效率。

时隙ALOHA要求网络中的所有节点同步，这一点会造成较大的开销。时隙化之后，如果多个用户试图在同一个时隙内发送分组时，碰撞也会发生。采用纠错编码可以使分组在碰撞后仍然能被正确接收，但若纠错能力不够强，这个碰撞的分组就必须重传。文献[16]研究了纠错编码和重传之间的最优设计问题。

例14.4：某时隙ALOHA系统的传输速率为R=10Mbit/s。假设每个分组有1000个比特。分组到达率λ为多少时，系统能达到最大的吞吐量？此时的有效数据速率是多少？

解：当$L=\lambda\tau=1$时，吞吐量达到最大，其中λ是分组到达率，τ是分组持续时间。对于

10Mbit/s的数据速率及1000比特的分组长度，分组持续时间是τ=1000/10^7=0.1ms，因此λ=1/0.0001=10^4，即每秒到达10^4个分组。L=1时的吞吐量是T=0.37，有效的数据速率为TR=3.7Mbit/s。与连续数据传输相比，数据率大约降低为1/3，降低的原因是分组到达的随机性和分组间的碰撞。

464

14.3.3 载波监听多址（CSMA）

减少碰撞的一种方法是采用载波监听多址技术，它让用户监听信道，如果发现其他用户正在发送，则将自己的分组推迟发送。检测出载波所需要的时间和传播时延都必须很小，否则会影响效率（见文献[17]的4.19节）。用户发现信道忙时，一般要等一段随机时间再发送。这种*随机退避*（random backoff）避免了信道变成空闲后，多个用户同时抢占信道的问题。采用载波监听多址协议需要用户能检测出其他用户是否在发送，并且传播时延也要足够小。有线局域网满足这些要求，因此以太网采用了CSMA。由于无线信道的特性，用户有可能检测不到其他用户的发送，这就是*隐藏终端问题*（hidden terminal problem）：每个节点只能监听到直接相邻的节点，不能检测出其他节点。如图14-7中，节点3和节点5都有数据要发给节点4，假设节点5已经在发送。由于距离太远，节点3检测不到节点5在发送，于是认为信道空闲，开始向节点4发送，结果是和节点5发送的信号发生碰撞。节点3不能检测出节点5的发送，它对节点5来说是隐藏的。

暴露终端（exposed terminal）问题也会降低采用CSMA的ALOHA系统的效率。图14-7中的节点2就是暴露节点。节点2有数据要发给节点1，与此同时节点3正在向节点4发送，节点2侦听信道时，发现信道已被占用，因此节点2就不会向节点1发送。但实际上，如果节点2向节点1发送的话是不会发生碰撞的。暴露终端只出现在多跳网络中，因此我们在第16章中再进行讨论。

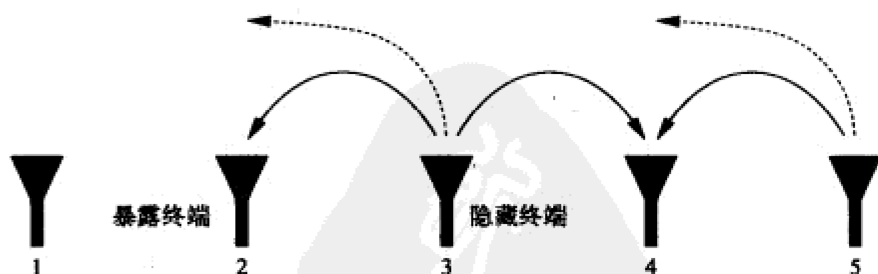


图14-7 隐藏终端和暴露终端

经常用四方握手（four-way handshake）的方法来避免隐藏节点所引起的碰撞[18-19]。*碰撞避免*（collision avoidance）的具体方法是：欲发送数据的节点先要等到信道空闲，然后发送一个很短的RTS（request to send）包，若目标接收节点允许传送，则会立即回应一个CTS（clear to send）包，CTS也同时告诉邻近的隐藏节点在它接收数据期间保持静默，这些隐藏节点在接收节点的通信范围内，但不在发送节点的通信范围内（对其隐藏）。其他检测到RTS或CTS的节点将在预期时间内保持静默。只有在信道空闲，

并且自己没有被其他RTS或CTS置为静默时，节点才可以发送RTS。只有在自己未被置为静默时，节点才能回应CTS。通常将RTS/CTS握手协议和随机退避结合运用，以避免信道变成空闲时，所有节点都开始发送。在一些具体标准[19~20]，如802.11 WLAN标准中（见文献[5]的14.3节），接收端在正确接收分组后发回一个ACK确认包，此后信道才被释放。 465

另外一种避免隐藏节点的方法是发送忙音。用户首先监听单独的控制信道以得知业务信道是否被占用（见文献[1]的4.6节），所谓"忙音"一般是指控制信道中的一个特定比特。如果网络设有一个能被所有用户监听到的中央控制器，那么这种方法可以有效避免碰撞。无中心的扁平网络则需要采取一些复杂的技术，以保证可能干扰业务信道的用户能够监听到控制信道上的忙音[21~22]。也可以将握手、忙音发送、功率控制等技术结合起来使用[22]。还可以在ALOHA中结合DSSS以减少碰撞。这种方案中的用户都使用相同的扩频码进行扩频调制，只要用户的发送时间能错开一个码片以上，碰撞引起的干扰就能被扩频码的自相关特性抑制[23]。

14.3.4 调度

随机接入协议非常适合于用户数多于可用信道数，并且每个用户实际发送时间很少的突发业务。如果用户要发送的是连续数据流，或者是一长串数据分组，那么大数据量将会加剧碰撞的发生，此时随机接入的性能就比较差。这种情况下提高性能的方法是采用调度技术对各用户的信道使用做出系统的安排。调度接入把可用带宽按时间、频率或码字分成信道，每个节点按某个时间表进行发送，其原则是避免相邻节点发生冲突，同时能充分利用信道资源。

调度也需要某种形式的ALOHA。作为一种随机接入方式，调度不可能在传输大量的数据之前预先排出时间表，这样的时间安排只能通过一些初始化过程完成，ALOHA的作用是完成初始化相关的信息交互。*分组预约多址*（packet-reservation access, PRMA）[24]就是这样的例子，它结合了突发数据的随机接入和连续数据的调度接入。PRMA假设在一个时隙化的系统中既有突发业务也有连续业务用户（如数据和语音用户），多个用户通过随机接入方式竞争某个时隙，如果某个用户在这个时隙上传输成功，以后这个时隙就是它的，直到它发完所有数据。如果用户要发送很长时间或者这个用户的业务本来就是连续的，这就等于给它分配了一个专用信道（假设后续的发送都是成功的，一旦由于信道原因使它有一次传错，它就失去了这个时隙信道，需要重新竞争，这样可能会造成很大的延时和丢包[25]）。用户发完后释放这个时隙，好让所有用户都能通过随机接入来竞争。这样，短业务的用户通过随机接入得到一个空闲时隙，连续业务的用户在初始竞争成功后，得到一个周期性时隙。文献[26]介绍了一种类似的结合预约和ALOHA的技术。 466

14.4 功率控制

功率控制用在用户间相互干扰的系统中，其目标是调整每个用户的发射功率，使所有用户的SINR均达到可接受性能所需要的门限值。不同的用户的门限值有可能不同，

PDG

取决于各自对性能的要求。下行链路中用户信号和干扰信号经历的信道增益相同，功率控制的问题很简单，但对于上行链路，由于信号和干扰经历的信道增益不同，问题就不是那么简单了。对蜂窝系统和自组织网络的功控研究最早见于文献[27~29]，上行功控是其特例。在上行模型中，第k个用户到接收机的信道增益为g_k，链路质量由SINR决定。对于有K个用户的上行信道，用户k的SINR为

$$\gamma_k = \frac{g_k P_k}{n + \rho \sum_{j \neq k} g_j P_j} \quad k = 1, \cdots, K \tag{14-5}$$

式中P_k是用户k的发射功率，n是接收端的噪声功率，ρ是信号处理所能带来的干扰衰减系数。例如CDMA的上行链路中，扩频码的处理增益能够降低干扰功率，因此，$\rho \approx 1/G$，G是处理增益，而在TDMA系统中$\rho=1$。

假设每条链路要求的最小SINR是$\gamma_k^* > 0$。这个约束条件可用矩阵形式表示为

$$(\boldsymbol{I}-\boldsymbol{F})\boldsymbol{P} \geqslant \boldsymbol{u} \quad \boldsymbol{P} > 0 \tag{14-6}$$

其中不等式逐元素成立，$\boldsymbol{P} = (P_1, P_2, \cdots, P_K)^T$为发射功率构成的向量，$\boldsymbol{u}$为

$$\boldsymbol{u} = \left(\frac{n\gamma_1^*}{g_1}, \frac{n\gamma_2^*}{g_2}, \cdots, \frac{n\gamma_K^*}{g_K}\right)^T \tag{14-7}$$

它是与信道增益和SINR要求成比例的噪声功率向量，$\boldsymbol{F}$是一个矩阵，其元素为

$$F_{kj} = \begin{cases} 0 & k = j \\ \gamma_k^* g_j \rho / g_k & k \neq j \end{cases} \tag{14-8}$$

其中$k, j = 1, 2, \cdots, K$。

矩阵$\boldsymbol{F}$的所有元素都是非负的，且$\boldsymbol{F}$不可约。令ρ_F为$\boldsymbol{F}$的Perron-Frobenius特征值，它是$\boldsymbol{F}$的特征值中模最大的。由于$\boldsymbol{F}$不可约，故ρ_F是正实数，且为$\boldsymbol{F}$的单特征值。根据Perron-Frobenius定理和矩阵论[30]，下列说法是等价的：

1. $\rho_F < 1$。
2. 存在向量$\boldsymbol{P} > 0$（即对于所有k有$P_k > 0$）使得$(\boldsymbol{I}-\boldsymbol{F})\boldsymbol{P} \geqslant \boldsymbol{u}$。
3. $(\boldsymbol{I}-\boldsymbol{F})^{-1}$存在且其元素均为正。

如果上述条件中任何一个成立，则$\boldsymbol{P}^* = (\boldsymbol{I}-\boldsymbol{F})^{-1}\boldsymbol{u}$是式（14-6）的Pareto最优解。就是说，满足式（14-6）的任何$\boldsymbol{P}$都逐项大于$\boldsymbol{P}^*$。因此$\boldsymbol{P}^*$是能够同时满足所有SINR要求的最佳功控，它使用户的发射功率最小。

467

文献[28]还证明，若$\rho_F < 1$，则下式给出的迭代Foschini-Miljanic功控算法能收敛到$\boldsymbol{P}^*$

$$\boldsymbol{P}(i+1) = \boldsymbol{F}\boldsymbol{P}(i) + \boldsymbol{u} \tag{14-9}$$

若$\rho_F \geqslant 1$，此算法趋于无穷大。这个算法可以简化为逐用户的形式：

$$P_k(i+1) = \frac{\gamma_k^*}{\gamma_k(i)} P_k(i) \tag{14-10}$$

其中$k \in \{1, 2, \cdots, N\}$。此即用户在SINR值低于目标值时增大发射功率，反之就减小发射

功率。SINR（或其函数，如BER）一般在基站或接入点测量，然后反馈一个简单的“增”“降”指令给各个发送端，以进行上述迭代。容易证明，式（14-9）和式（14-10）是逐项等价的，因此逐用户的功控算法也将收敛到$\mathbf{P}^*$。图14-8示出了两用户系统中能达到SINR目标值时用功率向量的允许区域，以及在此区域内可以收敛到最小功率向量$\mathbf{P}^*$的迭代算法。对于给定的一对目标SINR，图中的允许区域是所有能达到这个目标的功率对$\mathbf{P}=(P_1, P_2)$，最佳功率对$\mathbf{P}^*$是这个二维区域中功率最小的向量。

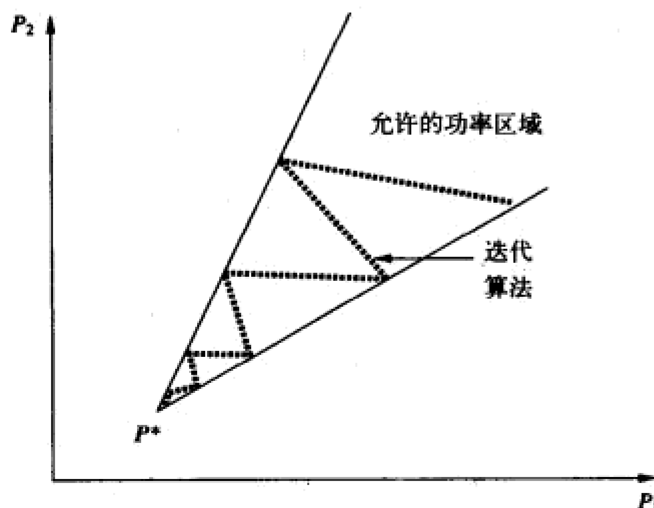


图14-8 迭代Foschini-Miljanic算法

Foschini-Miljanic功控算法也可以和接纳控制结合[31]。此时，是否允许接入要看新用户会不会使其他用户的SINR降到门限值以下。即当一个新用户请求接入时，基站或接入点需确定出是否存在一组发送功率，使得接纳了这个用户后还能满足其他用户的SINR要求。如果不能，则拒绝请求，反之则允许接入。如果是允许接入，则包括新用户在内的所有用户的功控算法将把发送功率调到一个新的允许功率向量上。 468

针对时延约束，文献[32]提出并分析了一种多址接入的功控策略。它通过动态规划的方法根据信道条件和时延约束优化发送功率。最佳功控有三种模式：信道很差并且允许的时延很大时，发射功率非常小；中等信道条件和时延下，发射功率较大；时延要求非常小时，发射功率非常大。和恒定发送功率相比，这种功控策略能大幅度节约发射功率，同时能满足业务的时延要求。

14.5 下行（广播）信道容量

多个用户共享一个信道时，信道容量不再由单独的一个数来描述。极端情形下，若多个用户中的某一个完全占用了所有信号维，则多用户信道容量退化为第4章中所讲的单用户信道容量。信道在用户之间的分法有无限多种，因此我们用*速率域*（rate region）来描述多用户信道的容量。速率域中的一个点是各个用户的传输速率组成的向量，以这些速率传输时，所有用户的误码率都可以任意小。各种多用户传输策略下的速率域的并

集称为多用户系统的容量域（capacity region）。容量域对于上行和下行是不同的，因为它们的信道模型有着本质的区别。不过，上下行信道的镜向对称特点也表明它们的容量有一定的关系。实际上，上下行信道之间有一种对偶性，使我们能够从一个信道的容量域推出另一个。在信道容量的分析中，常常把下行信道叫做广播信道（broadcast channel，BC），把上行信道叫做多址信道（multiple access channel，MAC[1]），我们也将采用这些术语。本节讨论广播信道的容量域，14.6节讨论多址信道的容量域，14.7节讨论广播信道和多址信道的对偶性，以及利用对偶性计算信道容量的方法。

在给出了在AWGN广播信道模型后，我们将讨论带串行干扰抵消的叠加码分（code division，CD）、时分（time division，TD）和频分（frequency division，FD）的方式叠加多个用户时的速率域，然后给出正交和非正交DSSS的速率域，同时我们还将讨论衰落条件下广播信道的容量问题。

我们将看到，达到容量的方法是采用干扰抵消的叠加码技术。采用串行干扰抵消的DSSS有一定的容量损失，扩频增益越大，容量损失也越大。此外，正交CD的扩频系统能够达到TD和FD容量域的子集，不采用干扰消除的非正交码分容量不如其他频谱共享技术。衰落条件下的容量域与收发是否已知信道信息有关，类似于单用户的情形。

469

14.5.1　信道模型

考虑一个广播信道中一个发射机向多个接收机发送多个不同的数据流，也称独立信息或独立数据。这与一般的广播电台或电视广播不同，那里是向不同的用户发送相同的数据流，也称公共信息或公共数据。不过如14.5.3节所述，这里的容量结论容易推广到公共信息的情形。广播信道的容量域反映可以同时发送给不同接收端的信息速率。我们主要考虑两用户的广播信道，由此得到的一般特性以及不同频谱共享技术的相对性能对两用户信道同样成立[33]。

在两用户广播信道中，一个发射机以速率R_k, $k=1, 2$向两个接收机发送数据，发射机到每个用户的信道增益为g_k, $k=1, 2$，每个接收机上的加性白高斯噪声的功率谱密度是$N_0/2$。定义第k个用户的有效噪声为$n_k=N_0/g_k$, $k=1, 2$。不妨假设$n_1\leqslant n_2$，即第一个用户的信道增益比第二个用户大。将噪声功率谱密度进行等效的做法并没有改变SINR，因为信号和干扰受到的信道衰减是一样的。因此信道增益为$\{g_k\}$的广播信道和噪声为$\{n_k\}$的广播信道容量相同[34]。对信道增益（或有效噪声）排序使其成为一个退化的广播信道（degraded broadcast channel），此时容量的一般表达式已知（见文献[35]的14.6节）。记总平均发射功率为P，总带宽为B。

如果将所有的功率和带宽都给一个用户，另一个用户的速率显然就是零，因此可达速率对(R_1, R_2)必然包括$(C_1, 0)$和$(C_2, 0)$，其中

$$C_k = B\log_2\left(1+\frac{P}{n_k B}\right) \quad k=1,2 \tag{14-11}$$

1. MAC也是medium access control（媒质接入层）的简写，见文献[1]的1.2节。

是4.1节给出的AWGN信道下单用户的信道容量，单位是每秒比特。这两个点就是广播信道容量域的边界。现在我们考虑容量域内的速率对，这个速率对可以用更公平一些的资源划分方法达到。

14.5.2 AWGN信道的容量

本节计算采用TD、FD以及可达到容量的最佳叠加码时，AWGN广播信道可实现的速率对。时分将发射功率P和带宽B在时间比例τ内分配给用户1，其余时间内分配给用户2。这种TD可达到$\mathcal{C}_1$和$\mathcal{C}_2$两点之间的连线，相应的速率对为：

$$\mathcal{C}_{TD}=\bigcup_{\{\tau:0<\tau<1\}}\left(R_1=\tau B\log_2\left(1+\frac{P}{n_1 B}\right)\quad R_2=(1-\tau)B\log_2\left(1+\frac{P}{n_2 B}\right)\right)\qquad(14\text{-}12)$$

470

速率域示于图14-10和图14-11中，其中n_1B和n_2B分别相差3dB和20dB。这个分贝差反映两个用户的信道增益差别，在比较不同复用方式的可实现速率时，这是一个重要的参数。

如果我们在保持平均功率仍为P的条件下调整各用户的发射功率，可实现的速率集能更大一些。假设P_1和P_2分别是两个用户的功率，则平均功率约束条件为$\tau P_1+(1-\tau)P_2=P$。于是，可变功率TD的可实现速率域为

$$\mathcal{C}_{TD,VP}=\bigcup_{\{\tau,P_1,P_2:0<\tau<1;\ \tau P_1+(1-\tau)P_2=P\}}\left(R_1=\tau B\log_2\left(1+\frac{P_1}{n_1 B}\right)\right.$$
$$\left.R_2=\left((1-\tau)B\log_2\left(1+\frac{P_2}{n_2 B}\right)\right)\right)\qquad(14\text{-}13)$$

在频分复用下，发送端分配把功率P_k和带宽B_k分配给第k个用户，功率和带宽约束为$P_1+P_2=P$和$B_1+B_2=B$。对于给定的频带划分(B_1,B_2)，可达到的速率对为

$$\mathcal{C}_{FFD}=\bigcup_{\{P_1,P_2:P_1+P_2=P\}}\left(R_1=B_1\log_2\left(1+\frac{P_1}{n_1 B_1}\right)\ R_2=B_2\log_2\left(1+\frac{P_2}{n_2 B_2}\right)\right)\qquad(14\text{-}14)$$

Bergmans和Cover在文献[33]中指出，固定分配带宽时，若n_1严格小于n_2，存在一个功率分配的范围$\{P_1,P_2:P_1+P_2=P\}$，使得相应的速率对能超出等功率TD的速率对（14-12）。这一点示于图14-10和图14-11中，图中画出了两种不同带宽分配下FD的速率域。图14-10中用户信道增益接近，超出不明显，但在图14-11中非常明显，此时用户的信道增益相差20dB。

FD的可达速率域定义为所有固定带宽分配下FD可达速率域的并

$$\mathcal{C}_{FD}=\bigcup_{\{P_1,P_2,B_1,B_2:P_1+P_2=P;\ B_1+B_2=B\}}\left(R_1=B_1\log_2\left(1+\frac{P_1}{n_1 B_1}\right)\right.$$
$$\left.R_2=B_2\log_2\left(1+\frac{P_2}{n_2 B_2}\right)\right)\qquad(14\text{-}15)$$

文献[33]指出，这个可达速率域超过了式（14-12）的等功率TD的速率域。这一点表现为图14-10和图14-11中几种固定带宽分配FD速率域的轮廓线，虽然在信噪比相近的图

14-10中不明显。实际上，$n_1 = n_2$时，式（14-15）退化为式（14-12）[33]。因此，最佳的功率分配和带宽分配在用户信道差别较大时更为有益。

471

注意不等功率TD的可达速率域（14-13）和式（14-15）中的FD速率域是相同的。在式（14-13）中令$B_i = \tau_i B$、$\pi_i = \tau_i P_i$，其中$\tau_1 = \tau$、$\tau_2 = 1-\tau$，则功率约束条件成为$\pi_1 + \pi_2 = P$。代入式（14-13）得到

$$\mathcal{C}_{\mathrm{TD,VP}} = \bigcup_{\{\pi_1,\pi_2:\pi_1+\pi_2=P\}} \left(R_1 = B_1 \log_2\left(1+\frac{\pi_1}{n_1 B_1}\right) \quad R_2 = B_2 \log_2\left(1+\frac{\pi_2}{n_2 B_2}\right) \right) \tag{14-16}$$

与式（14-14）比较可见，只要取适当的P_i和τ_i，FD速率域上的任何点都能通过可变功率TD的时分达到。

带有干扰消除的叠加码是一种多分辨率编码技术。在叠加码中，信道增益高的用户能细分星座，信道增益低的用户只能粗分星座（见文献[33]及文献[35]的14.6节）。例如文献[36]给出的一种两级叠加码采的星座是嵌入了4PSK的32QAM，如图14-9所示。发送端发送的是32QAM星座点，星座点的选取方式是：先用信噪比较差的用户（即用户2，因为$n_2>n_1$）的两个比特选择出图中4PSK叠加点中的一个，再用信噪比较好的用户的三个比特选择出这个4PSK星座点周围的八个星座点中的一个。到达接收端时，信道条件好的用户能容易地分辨出星座点所在的象限，从而减去4PSK叠加点；信道条件差的用户不能区分出4PSK点周围的星座点，对他来说叠加在4PSK上的32QAM如同噪声，他只能解调出4PSK星座。采用更复杂的星座很容易把这种思想推广到多用户的情形。叠加码通过扩展星座的方式实现了多速率，不需要增加带宽。

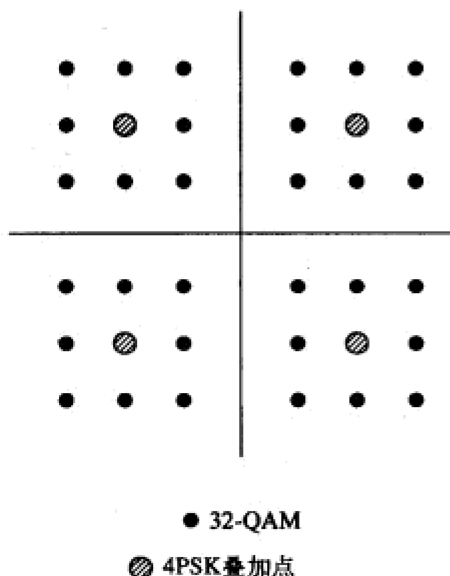


图14-9　嵌入4PSK的32-QAM

对于采用叠加码和串行干扰抵消的两用户系统，文献[33]给出容量域为如下的速率对集合：

$$\mathcal{C}_{\mathrm{BC}} = \bigcup_{\{P_1,P_2:P_1+P_2=P\}} \left(R_1 = B \log_2\left(1+\frac{P_1}{n_1 B}\right) \quad R_2 = B \log_2\left(1+\frac{P_2}{n_2 B + P_1}\right) \right) \tag{14-17}$$

对上式的直观解释也和图14-9的例子一样。由于$n_1<n_2$，用户1能正确接收发给用户2的信息并减去其影响，然后译出发送给用户1自己的信息。用户2的信噪比差，不能正确译出用户1的信息，于是功率为P_1的用户1的信号对用户2来说就是干扰。达到容量时每个用户发送的信号都服从高斯分布（见文献[33]及文献[35]的14.1节），故可将这个干扰看成是加性高斯白噪声，13.4.4节中DSSS的串行干扰抵消也是这样处理的。虽然串行干扰抵消能达到式（14-17）中的容量域，但它在实际应用中不一定是最好的方法。容量分析

中假设了理想的译码和理想的信道估计，而实际的译码器总会有一些误码，信道估计也有一定误差，这将使串行干扰抵消中的判决反馈出现错误。因此在实际应用中，不受这些因素影响的多用户检测方法有可能比串行干扰抵消有更好的性能。 472

当$n_1<n_2$时，文献[37]证明式（14-17）所给出的速率域超过了TD和FD可达到的速率域。文献[37]还证明，这个速率域是任何形式的编码和频谱共享技术所能达到的最大速率域，因此式（14-17）就是广播信道的容量域，记为C_{BC}。如果所有用户有相同的信噪比，此容量域将退化成等功率TD的直线（14-12）。因此当$n_1 = n_2$时，任何频谱共享方式都有相同的速率域。

叠加码的思想容易推广到用户数$K>2$的多用户系统中。对于有K个用户的广播信道，假设每个用户的信道增益为g_k。把用户按其有效噪声$n_k = N_0/g_k$排序，按这个次序设计的叠加码有K级，有效噪声最大（即信噪比最差）的用户只能译出最粗的一级，噪声次大的能译出下一级，依次类推。每个用户都能消除比自己噪声大的用户的星座点的影响，但比自己信道好的用户的星座点表现为噪声。设总功率约束为P，则式（14-17）推广到多用户时为

$$C_{\mathrm{BC}} = \bigcup_{\{P_k:\sum_{k=1}^{K}P_k=P\}}\left\{(R_1,\cdots,R_K): R_k = B\log_2\left(1+\frac{P_k}{n_kB+\sum_{j=1}^{K}P_j\mathbf{1}[n_k>n_j]}\right)\right\} \tag{14-18}$$

式中$\mathbf{1}[.]$为示性函数。

广播信道的*和速率容量*（sum-rate capacity）是容量域中所有速率矢量的速率和的最大值：

$$C_{\mathrm{BCSR}} = \max_{(R_1,\cdots,R_K)\in C_{\mathrm{BC}}}\sum_{k=1}^{K}R_k \tag{14-19}$$

和速率容量是一个标量，它是系统在不考虑用户速率分配的公平性时能达到的最大吞吐量。它比K维的容量域简单，并且也能反映出一些重要的特性。对于AWGN广播信道，由式（14-18）可以证明，将所有功率分配给具有最大信道增益（也即有效噪声最小）的用户时，和速率容量可达到最大。就是说K用户AWGN广播信道的和速率容量C_{BCSR}为

$$C_{\mathrm{BCSR}} = B\log_2\left(1+\frac{P}{n_{\min}B}\right) = B\log_2\left(1+\frac{g_{\max}P}{N_0B}\right) \tag{14-20}$$

其中$n_{\min}\triangleq\min_k n_k$，并且$g_{\max}\triangleq\max_k g_k$。和速率容量点就是式（14-11）中容量域的一个边界点。由于所有的系统资源都分配给了一个用户，所以C_{BCSR}对叠加码、时分和频分都是相同的。 473

例14.5：假设AWGN广播信道的总发射功率为P=10mW，$n_1=10^{-9}$W/Hz，$n_2=10^{-8}$W/Hz，带宽B=100kHz。若用户1所需速率为300kbit/s，求等功率TD、等带宽FD和叠加码下可分配给用户2的速率。

解：等功率TD中用户1的速率为$R_1=\tau B\log_2(1+P/n_1B) = (6.658\times10^5)\tau$bit/s。令其等于所要

求的速率：$R_1=(6.658\times 10^5)\tau = 3\times 10^5$bit/s，得到$\tau = 3\times 10^5/(6.644\times 10^5) = 0.451$。因此用户2可得到的速率为$R_2 = (1-\tau)B\log_2(1+P/n_2B) = 1.89\times 10^5$bit/s。等带宽FD时，用户1的速率为$R_1 = 0.5B\log_2(1+P_1/(0.5\times n_1B)) = 3\times 10^5$bit/s，由此得到$P_1 = 0.5\times n_1B(2^{R_1/0.5B}-1)$=3.15mW，因此$P_2 = P-P_1 = 6.85$mW，$R_2 = 0.5B\log_2(1+P_2/(0.5\times n_2B)) = 1.94\times 10^5$bit/s。叠加码中，$R_1 = B\log_2(1+P_1/n_1B) = 3\times 10^5$bit/s，解得$P_1 = n_1B(2^{R_1/B}-1) = 0.7$mW，因此$R_2 = B\log_2\left(1+\dfrac{P-P_1}{n_2B+P_1}\right) = 2.69\times 10^5$bit/s。可以看出叠加码明显优于TD和FD。对TD进行功率分配优化，对FD进行带宽分配优化后这个差距可以缩小。

例14.6：给出前例所示系统的和速率容量。

解：已知条件是P=10mW、$n_1=10^{-9}$W/Hz、$n_2=10^{-8}$W/Hz、B=100kHz。用户1的噪声最小为$n_{\min} = 10^{-9}$W/Hz，因此$C_{\mathrm{BCSR}} = B\log_2(1+P/n_{\min}B) = 6.644\times 10^5$bit/s。由于是把所有的资源都分配给一个用户，所以叠加码、TD及FD和速率容量都能达到这个和速率容量。

13.4节已经指出，多用户码分也可以通过DSSS来实现。这样的系统用不同的扩频码对各个用户的信号进行调制，发送信号的带宽将大约增加G倍，其中G是扩频码的处理增益。正交扩频码的互相关为零，要得到N个扩频码，相应得扩频码增益必须是N。若系统带宽受限为B，则每个用户的信息带宽就是B/N。于是，正交扩频的两用户速率域为

$$C_{\mathrm{DS,OC}} = \bigcup_{\{P_1,P_2:P_1+P_2=P\}}\left(R_1 = \frac{B}{2}\log_2\left(1+\frac{P_1}{n_1B/2}\right)\qquad R_2=\frac{B}{2}\log_2\left(1+\frac{P_2}{n_2B/2}\right)\right) \tag{14-21}$$

474

比较式（14-21）和式（14-14）可见，正交扩频的两用户DSSS的容量域和等带宽分配的频分($B_1 = B_2 = B/2$)系统相同。与式（14-16）对比，不等功率TD也能达到同样的速率域。因此，采用正交码（如Walsh-Hadamard码）的DSSS系统可达到的速率对是FD和TD速率域的子集，如欲相同，则需要采用更一般化的正交码。

再来考虑非正交扩频码的DSSS系统。如13.4.2节所论，这些系统中的用户间干扰被扩频码的互相关性衰减。若将用户间的干扰看成噪声，那么干扰功率对SIR的贡献降低为互相关的平方。根据式（13-6），可假设处理增益为G的扩频码能将干扰功率衰减为原来的$1/G$。这对于随机码是一个合理的近似，但对干扰功率的衰减的实际值与具体的扩频码特性及其他假设有关（见13.4.1节及文献[38~39]）。由于带宽增加了G倍，因此，带串行干扰抵消的两用户非正交DSSS系统的速率域为

$$C_{\mathrm{DS,NC,IC}} \bigcup_{\{P_1,P_2:P_1+P_2=P\}}\left(R_1 = \frac{B}{G}\log_2\left(1+\frac{P_1}{n_1B/G}\right)\right.$$
$$\left. R_2 = \frac{B}{G}\log_2\left(1+\frac{P_2}{n_2B/G+P_1/G}\right)\right) \tag{14-22}$$

由log函数的上凸性可知，$G>1$时上式的速率域要比式（14-17）的叠加码小，且G越大，差距也越大。说明，如欲使非正交DSSS的容量最大化，必须要使扩频增益最小化。

对于无干扰抵消的非正交码扩频，接收端把其他用户的信号都当作噪声，于是速率域为

$$C_{\mathrm{DS,NC}} = \bigcup_{\{P_1,P_2:P_1+P_2=P\}} \left(R_1 = \frac{B}{G}\log_2\left(1+\frac{P_1}{n_1B/G+P_2/G}\right)\right.$$
$$\left. R_2 = \frac{B}{G}\log_2\left(1+\frac{P_2}{n_2B/G+P_1/G}\right)\right) \quad (14\text{-}23)$$

同样由log函数的上凸性可知，$G=1$时速率域达到最大，G越大速率域越小。此外，式（14-23）的曲率半径为

$$\chi = \frac{\dot{R}_1\ddot{R}_2 - \dot{R}_2\ddot{R}_1}{(\dot{R}_1^2+\dot{R}_2^2)^{3/2}} \quad (14\text{-}24)$$

式中$\dot{R}_i$和$\ddot{R}_i$分别是R_i对于α的一阶导和二阶导，$P_1=\alpha P$、$P_2=(1-\alpha)P$。$G=1$时$\chi\geqslant 0$，表明无干扰抵消非正交扩频的速率域（14-23）是以C_1和C_2为端点的凹函数，如图14-10和图14-11所示。因此，无干扰抵消的非正交DSSS的速率域小于频分和时分的速率域，它们的速率域是相同端点的凸函数。 475

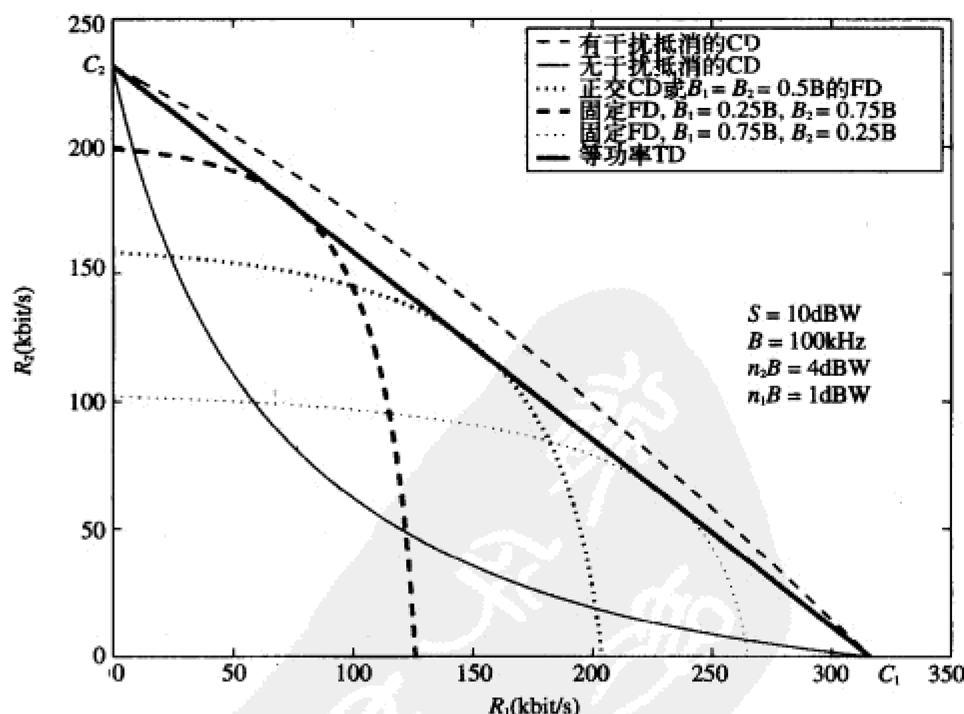


图14-10 两用户信噪比相差3dB时的容量域

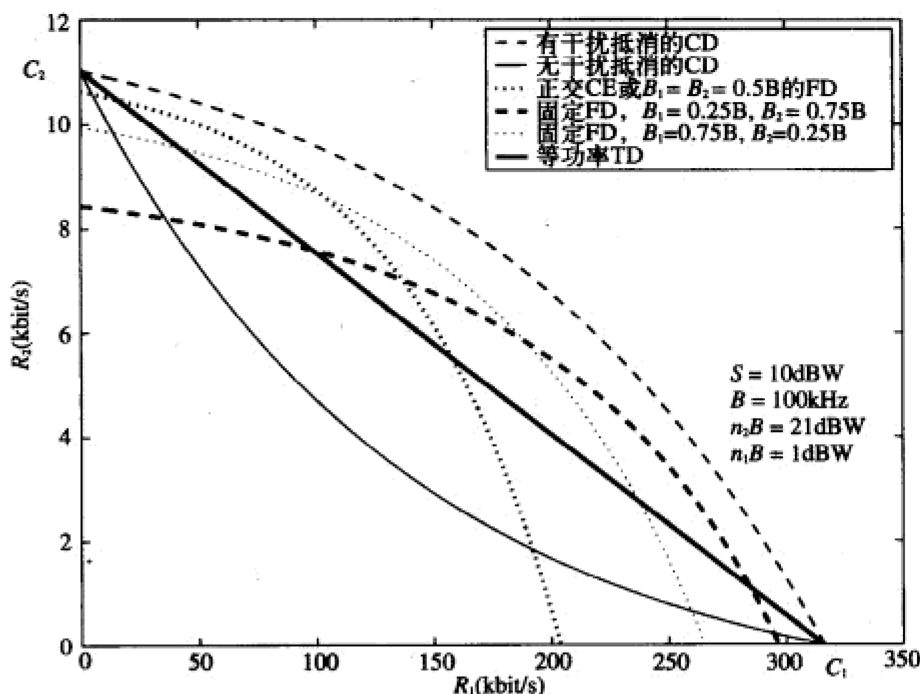


图14-11 两用户信噪比相差20dB时的容量域

图14-10和图14-11分别画出了两用户信噪比相差3dB和20dB时不同情形下的速率域，包括等功率TD（14-12）、FD（14-14）、正交DSSS（14-21）、有干扰抵消的非正交DSSS（14-17）、无干扰抵消的非正交DSSS（14-23）。计算式（14-23）时假设了$G=1$，若直序扩频的扩频增益较大，速率域将较小。

14.5.3 公共数据

在许多广播业务中，系统发送公共数据给所有用户。例如电视台和广播电台就是给所有的用户发送相同的数据。在无线Internet应用中，许多用户也可能下载相同的股票信息和体育比赛成绩。叠加码的特性使它能直接成为广播公共数据的最优技术，并能使广播信道的容量域纳入公共数据。例如对于采用叠加码的两用户广播信道，信道好的用户除了能接收到发给自己的数据外，还能收到发给信道差的用户的数据。公共数据必须要同时发给两个用户，我们可以把公共数据当作发送给信噪比差的用户的独立数据来进行编码发送，这样，两个用户都能收到这个数据。

476

在这种发送策略下，若(R_1, R_2)是发送独立数据时两用户广播信道的容量域（14-17），那么对于任意的$R_0 \leqslant R_2$，该广播信道可以达到速率组(R_0, R_1, R_2-R_0)，其中R_0是公共数据速率、R_1是用户1的独立数据速率、R_2-R_0是用户2的独立数据速率。这样的三维容量域可以在数学上表示为

$$C_{BC}=\bigcup_{\{P_1,P_2:P_1+P_2=P\}}\left(R_0\leqslant B\log_2\left(1+\frac{P_2}{n_2B+P_1}\right)\ R_1=B\log_2\left(1+\frac{P_1}{n_1B}\right)\right.$$

$$\left.R_2=B\log_2\left(1+\frac{P_2}{n_2B+P_1}\right)-R_0\right) \quad (14\text{-}25)$$

例14.7：从例14.5可知，假设广播信道的总功率为P=10mW、n_1= 10^{-9}W/Hz、n_2= 10^{-8}W/Hz、B=100kHz，那么速率对$(R_1, R_2) = (3\times10^5, 2.69\times10^5)$在容量域的边界上。若用户1要求的独立数据速率为300kbit/s，系统要求公共数据速率为100kbit/s，那么用户2能得到的独立数据速率是多少？

解：为使用户1的独立数据速率为R_1=300kbit/s，与例14.5.2中相同，需要P_1=0.7mW。公共信息数据速率R_0= $10^5<2.69\times10^5$，由式（14-25）可得用户2的独立数据速率为R_2-R_0= $2.69\times10^5-10^5=1.69\times10^5$kbit/s。

477

14.5.4 衰落信道容量

下面讨论存在衰落的广播信道的容量域，此时各用户的信道增益是独立时变的随机值。如在4.2节中所讨论的单用户衰落信道容量，衰落广播信道的容量也与发送端和接收端对信道的了解程度有关。然而，目前只清楚退化广播信道的容量，而这个模型假设收发两端都已知信道信息。此外，如果发送端不知道信道信息的话，也无法使用叠加码，因为不知道哪个用户能粗分星座点，哪个能细分星座点。因此我们只考虑这样的广播信道，它的收发两端都理想已知关于瞬时信道增益的信道边信息（channel side information，CSI）。同时假设信道衰落的变化足够慢，使得对于容量域内的任何点，所用的编码技术能在信道状态改变之前使错误率趋于零[1]。

与单用户衰落信道类似，理想CSI下多用户衰落信道的容量也有遍历容量（香农容量）和中断容量之分。多用户衰落信道的遍历容量域是对衰落平均的可实现的速率向量的集合[34][40]；中断容量域是除去一定中断率之外的衰落状态外，其余衰落状态下能够达到的固定速率向量的集合[41-43]。零中断容量域指中断率为零时的中断容量域，即所有衰落状态下都可以达到的固定速率向量的集合。遍历容量域对应单用户的遍历容量，表示不考虑时延限制时能达到的数据速率向量，某些衰落状态下的数据速率向量可能很小或为零。这对话音、视频这类有时延限制的业务来说会是一个问题。中断容量域对应单用户的中断容量，它对于所有非中断的衰落状态使用同一个固定的速率向量，这个容量指标对时延有限制的业务可能比较适合。由于中断容量域要求在某些深衰落下也能维持同样的固定速率，所以它也将大大低于遍历容量域。实际上，当所有用户都是瑞历衰落时，零中断容量域是零。

1. 确切地说，能达到AWGN广播信道容量域中的一个点的编码采用了分组码，其错误率随码长的增加而趋于零。我们的慢衰落假设就是说，信道增益在足够长的时间内是常数，使得针对这个信道增益所设计的分组码的错误率可趋近零。

考虑一个存在加性高斯白噪声和衰落的广播信道，单个发射机以平均功率 $\bar{P}$ 在带宽B上向K个接收端发送独立信息。发送端和接收端都只有一个天线，用户k在i时刻的功率增益为$g_k[i]$。每个接收机的加性高斯白噪声的功率谱密度为$N_0/2$，用户K的有效噪声为$n_k[i]=N_0/g_k[i]$，i时刻的有效噪声向量[1]为

478

$$\boldsymbol{n}[i]=(n_1[i],\cdots,n_K[i]) \tag{14-26}$$

由于上式也反映了每个用户的信道增益，也称其为i时刻的衰落状态。在时间参考清晰的条件下，我们也会把$n_k[i]$记为n_k。如同无衰落信道，衰落广播信道的容量可以用时变信道增益或时变噪声向量来计算。广播信道的遍历容量域是衰落信道下能以任意小的错误率传输的平均速率向量的集合，平均是对衰落状态进行。文献[34]将衰落信道分解为一组并行的静态广播信道，每一个并行信道对应一个可能的信道状态$n=(N_0/g_1,\cdots,N_0/g_k)$，从而得到了衰落广播信道的遍历容量域和最佳的功率分配方案。每一个衰落状态可以看成是静态的AWGN广播信道，可以采用时分、频分或码分技术。

由于收发端都已知有效噪声向量$\boldsymbol{n}[i]$，因此发送端可以按照当前的噪声大小顺序来使用叠加码。每个接收机都能进行串行译码，先译出噪声比自己大的用户的信息，减掉后再译出自己的信息。此外，发送给每个用户的功率都是当前衰落状态的函数$P_j(n)$。发送方案是基于叠加码的，需要确定的是对用户和对时间的最佳功率分配。

定义衰落状态下的功率策略$\mathcal{P}$为从衰落状态n到各用户的发射功率$P_k(n)$的映射。令$\mathcal{F}_{BC}$表示满足功率约束$\bar{P}$的所有功率策略的集合：

$$\mathcal{F}_{\mathrm{BC}}\triangleq\left\{\mathcal{P}:E_{\boldsymbol{n}}\left[\sum_{k=1}^{K}P_k(\boldsymbol{n})\right]\leqslant\bar{P}\right\} \tag{14-27}$$

对于固定的衰落状态n及功率分配$P(n)=\{P_k(n): k=1,\cdots,K\}$，由式（14-18）得到容量域为

$$\mathcal{C}_{\mathrm{BC}}(\boldsymbol{P}(\boldsymbol{n}))=\left\{(R_1(\boldsymbol{P}(\boldsymbol{n}),\cdots,R_K(\boldsymbol{P}(\boldsymbol{n})):\right.$$
$$\left.R_k(\boldsymbol{P}(\boldsymbol{n}))=B\log_2\left(1+\frac{P_k(\boldsymbol{n})}{n_kB+\sum_{j=1}^{K}P_j(\boldsymbol{n})1[n_k>n_j]}\right)\right\} \tag{14-28}$$

令$\mathcal{C}_{BC}(\mathcal{P})$表示功控策略$P$下，对所有衰落状态取平均的可达速率集合：

$$\mathcal{C}_{\mathrm{BC}}(\mathcal{P})=\{(R_1,\cdots,R_K):R_k\leqslant E_{\boldsymbol{n}}[R_k(\boldsymbol{P}(\boldsymbol{n}))]\}$$

其中的$R_K(P(n))$由式（14-28）给出。根据文献[34]，在平均功率约束为$\bar{P}$时，有理想CSI的广播信道的遍历容量域为：

$$\mathcal{C}_{\mathrm{BC}}(\bar{P})=\bigcup_{\mathcal{P}\in\mathcal{F}_{\mathrm{BC}}}\mathcal{C}_{\mathrm{BC}}(\mathcal{P}) \tag{14-29}$$

479

文献[34]中还证明$\mathcal{C}_{BC}(\bar{P})$是上凸的，并且最佳功控是有$K$个注水线的一种扩展的注水法。

---

1. 注意噪声向量是噪声的瞬时功率，不是噪声的瞬时取样。

也可以定义出TD和FD的可达速率向量，其速率域显然要小于遍历容量域，因为在每个衰落状态下叠加码都优于它们。最佳TD根据当前衰落状态优化各用户的功率分配，最佳FD根据当前的衰落状态优化各用户的功率分配和带宽分配。根据14.5.2节，在每个衰落状态下，可变功率TD和功率带宽都可变的FD速率域相同，因此对所有衰落平均后，TD和FD的平均速率域也相同。考虑FD，对于衰落状态n；功率分配$\mathcal{P}\in\mathcal{F}_{BC}$给用户$k$分配的功率是$P_k(n)$，根据式（14-27），$P$满足平均功率约束条件。假设带宽分配$\mathcal{B}$分配给用户$k$的带宽为$B_k(n)$。令$\mathcal{G}$表示所有满足系统带宽约束条件的带宽分配的集合，即

$$\mathcal{G}\triangleq\left\{\mathcal{B}:\sum_{k=1}^{K}B_k(n)=B\ \forall n\right\}$$

在功率分配$\mathcal{P}$和带宽分配$\mathcal{B}$下，FD的可达速率域为

$$\mathcal{C}_{\text{FD}}(\mathcal{P},\mathcal{B})=\{(R_1,\cdots,R_K):R_k\leqslant E_n[R_k(P(n),\mathcal{B})]\}\tag{14-30}$$

其中

$$R_k(P(n),\mathcal{B})=B_k(n)\log_2\left(1+\frac{P_k(n)}{n_kB_k(n)}\right)\tag{14-31}$$

因此在带宽约束为B，功率约束为$\overline{P}$的条件下，有理想CSI的FD的可达速率域的集合为

$$\mathcal{C}_{\text{FD}}(\bar{P},B)=\bigcup_{\mathcal{P}\in\mathcal{F}_{\text{BC}},\,\mathcal{B}\in\mathcal{G}}\mathcal{C}_{\text{FD}}(\mathcal{P},\mathcal{B})\tag{14-32}$$

在衰落广播信道的遍历容量域中，速率和的最大值就是该信道的和速率容量。在AWGN广播信道中，发送端只对信道条件最好的用户发送，可以使和速率容量达到最大。对应到衰落广播信道中就是，发送端在每一信道状态下选择信道条件最好的用户发送，就能使和速率容量达到最大。由于每个衰落状态下资源仅分配给一个用户，所以CD、TD和FD三种方法的和速率容量都是相同的。计算和速率容量以及沿时间的最佳功率分配时，可以采用一个等效的单用户衰落信道，其平均功率约束为$\overline{P}$，时变有效噪声为$n[i]=\min_k n_k[i]$。根据4.2.4节的结论，在i时刻对信道最好的用户的功率分配是时间上的注水法，截止门限由$\min_k n_k[i]$的分布决定。

文献[34]在不同的衰落分布下给出了衰落广播信道的遍历容量以及CD、TD和FD下的可达速率域，以及可达到域边界的最佳的自适应资源分配方法。这些自适应策略利用多用户分集（multiuser diversity）将更多的资源（功率、带宽、时隙）分配给任一衰落状态下信道条件最好的用户。当我们把所有资源都分配给信道条件最好的用户时，就达到了和速率容量。14.8节将具体讨论多用户分集。 480

广播信道的零中断容量是在平均功率受限条件下，所有用户在所有衰落状态下都能达到的速率集合，它是4.2.4节中单用户信道零中断容量的多用户扩展。根据文献[42]，对于衰落状态n，达到速率向量$\boldsymbol{R}=(R_1,R_2,L,R_K)$所需要的功率为：

$$P^{\min}(\boldsymbol{R},\boldsymbol{n})=\sum_{k=1}^{K-1}\left[2^{(\sum_{j=k+1}^{K}R_{\pi(j)}/B)}(2^{R_{\pi(k)}/B}-1)n_{\pi(k)}B\right]+(2^{R_{\pi(K)}/B}-1)n_{\pi(K)}B\tag{14-33}$$

式中π(·)表示排序，以使

$$n_{\pi(1)} < n_{\pi(2)} < \cdots < n_{\pi(K)}$$

因此，零中断容量就是所有能满足平均功率约束条件的速率向量的集合：

$$\mathcal{C}_{\mathrm{BC}}^{0}(\bar{P}) = \bigcup_{\{R:E_n[P^{\min}(R,n)]\le\bar{P}\}} R = (R_1, R_2, \cdots, R_K) \tag{14-34}$$

零中断容量的边界就是式中平均功率约束取等式时的速率向量R的集合。对于两用户广播信道，若时变加性高斯白噪声的功率分别为n_1、n_2，则零中断容量的边界简化为满足下式的速率向量(R_1, R_2)的集合[42]：

$$\begin{aligned}\bar{P} = {} & p(n_1 < n_2)\left[E[n_1B \mid n_1 < n_2]2^{R_2/B}(2^{R_1/B}-1) + E[n_2B \mid n_1 < n_2](2^{R_2/B}-1)\right] \\ & + p(n_1 \geqslant n_2)\left[E[n_2B \mid n_1 \geqslant n_2]2^{R_1/B}(2^{R_2/B}-1) + E[n_1B \mid n_1 \geqslant n_2](2^{R_1/B}-1)\right]\end{aligned}$$

边界完全由$E[n_1B|n_1<n_2]$、$E[n_2B|n_1<n_2]$、$E[n_2B|n_1\geqslant n_2]$、$E[n_1B|n_1\geqslant n_2]$决定。这是因为从式（14-33）可见，在每一衰落状态下，达到速率向量时所需要的功率是噪声功率的线性函数。对噪声功率的不同排序将使式（14-33）不同，这一点使零中断容量取决于噪声功率的条件期望，而不是噪声功率的期望。

广播信道中断容量域的定义与零中断容量域的定义类似，只不过现在允许用户的中断率不为零，它们在中断状态可以停止传输。这使系统增加了灵活性，因为当衰落严重的情况下，要在所有衰落状态下都能维持固定的速率会耗费大量的功率。例如在4.2.4节的单用户瑞利衰落信道中，要维持一个固定的非零速率需要无限大的功率。正是因为允许某些中断，功率可以由于中断状态保留下来，从而维持了在非零中断状态下的高速率。中断容量域比零中断容量域更难以计算，困难在于对任意一个衰落状态，发送端需要确定出中断哪个用户。确定出了中断用户后，对剩余的$K' \leqslant K$个用户可用式（14-33）计算出维持它们的速率向量所需要的功率。文献[42]指出应通过门限策略来判定中断那个用户，然后可根据门限策略和式（14-33）的功率分配间接得到中断容量域。

481

将遍历容量和中断容量的概念结合起来便是最小速率容量域[44]。该域中的速率向量表示所有衰落状态下都大于某个最小速率向量的平均速率（对衰落平均），或者也可以允许有一定的中断率。最小速率容量对同时支持时延受限和时延不受限业务的系统很有用，最小速率是时延受限业务在所有衰落状态下可以保持的速率，在此之上的速率是时延不受限业务的可用速率，它随当前衰落的状态而变。零中断率时的最小速率容量域处于零中断容量域和遍历容量域之间：若最小速率为0，它就是遍历容量域；若最小速率是零中断容量域边界上的速率，它就不能超越这些边界点。这一点示于图14-12中，图中画出了广播信道在莱斯衰落下的遍历容量域、零中断容量域和最小速率容量域。从图中我们可以看到，遍历容量域最大，因为它根据不同的信道状态进行自适应以最大化平均速率（对衰落平均）。零中断容量域最小，因为它要求所有状态下都必须维持一个固定的速率，这在严重衰落的情况下会耗费大量的功率。最小速率容量域处于两者之间，具体取决于最小速率的要求。当所有衰落状态下都必须要维持的这个最小速率向量增加时，最小速率容量域将趋向于零中断容量域，反之则趋向于遍历容量域。

482

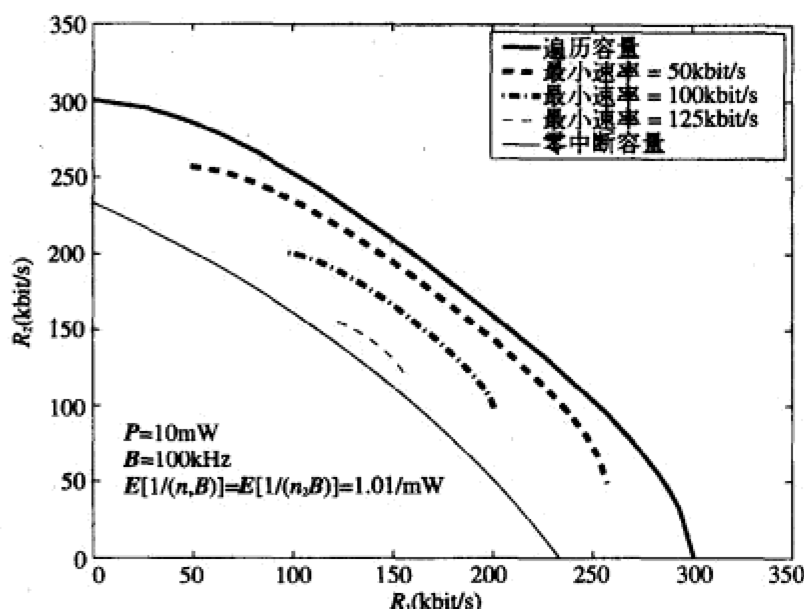


图14-12 广播信道的遍历容量域、零中断容量域和最小速率容量域
（K因子为1的莱斯信道，平均信噪比10dB）

14.5.5 多天线容量

现在讨论多天线广播信道的容量域。从10.3节中可以知道，多入多出（multiple-input multiple-output，MIMO）能大大提高单用户系统的容量。多用户系统也如此，实际上多用户系统更能有效地利用多个空间维。

考虑有K个用户的广播信道，发送端有M_t根天线，每个接收端都有M_r根天线。$M_r \times M_t$的信道矩阵$\boldsymbol{H}_k$表示发送端的每个天线到用户k的接收端的每个天线之间的信道增益。用户k的接收信号为

$$\boldsymbol{y}_k = \boldsymbol{H}_k\boldsymbol{x} + \boldsymbol{n}_k \tag{14-35}$$

式中x是发射天线的输入，其协方差矩阵为$\boldsymbol{Q}_x = E[\boldsymbol{x}\boldsymbol{x}^H]$。简便起见，我们把带宽归一化[1]为$B$=1Hz，并假设噪声向量$\boldsymbol{n}_k$为循环对称复高斯，$\boldsymbol{n}_k \sim N(0, \boldsymbol{I})$。

当发送端有多个天线（$M_t > 1$）时，广播信道不再是退化的，就是说一般不能把接收端按其信道质量排出次序，因为对应发送端不同的天线，接收端有多个信道增益。一般未退化广播信道的容量域还是未知的，不过文献[46~47]给出了它的可达速率域，后来文献[48]证明它就是容量域。这个域基于*脏纸编码*（dirty paper coding，DPC）[49]的概念，其基本思想是：对于某个特定的接收端，如果发送端理想已知非因果的干扰信息，则信道容量和无干扰时一样，也即和接收端完全已知干扰信息并能将干扰完全减掉时一样。脏纸编码能够在发送端预先消除非因果已知的干扰，同时又不会增加发射功率。关

1. 单位带宽下MIMO实（一维）信道的容量在log前有系数0.5，复（二维）信道无此系数（见文献[45]的3.1节）。

于这种预干扰抵消的实际技术及一般情形可参见文献[50]。

在MIMO广播信道中，发送端可以用DPC来选择不同用户的码字。发送端先选出用户1的码字x_1，再根据完全已知的用户1码字的信息选出用户2的码字x_2，因此用户1的码字可以预先被减去而使用户2在接收端看不见用户1的干扰。类似地，用户3的码字x_3也如此选出，使得用户3在接收端看不到用户1、2的干扰。依此类推，直至K个用户都完成编码。用户的顺序在这个过程中明显非常重要，计算容量时需要进行优化。令π(·) 表示用户标号的排列，$\boldsymbol{Q}=[\boldsymbol{Q}_1,\cdots,\boldsymbol{Q}_K]$表示一组满足$\mathrm{Tr}[\boldsymbol{Q}_1+\cdots+\boldsymbol{Q}_K]\leqslant P$的半正定协方差矩阵。若DPC是按π(1)、π(2)、…、的次序编码，则如下的速率向量是可达到的：

483

$$R_{\pi(k)}=\log_2\frac{\det[\boldsymbol{I}+\boldsymbol{H}_{\pi(k)}(\sum_{j\geqslant k}\boldsymbol{Q}_{\pi(j)})\boldsymbol{H}_{\pi(k)}^H]}{\det[\boldsymbol{I}+\boldsymbol{H}_{\pi(k)}(\sum_{j>k}\boldsymbol{Q}_{\pi(j)})\boldsymbol{H}_{\pi(k)}^H]}\quad k=1,\cdots,K \tag{14-36}$$

对于所有的用户次序排列、所有的满足平均功率受限的半正定协方差矩阵，速率向量集合的凸包就是容量域$\mathcal{C}$：

$$\mathcal{C}_{\mathrm{BC}}(P,\boldsymbol{H})\triangleq\mathrm{Co}\left(\bigcup_{\pi,\boldsymbol{Q}}\boldsymbol{R}(\pi,\boldsymbol{Q})\right) \tag{14-37}$$

其中$\boldsymbol{R}(\pi,\boldsymbol{Q})$由式（14-36）给出。发送信号是$x=x_1+\cdots+x_K$，输入协方差矩阵为$\boldsymbol{Q}_k=\boldsymbol{E}[\boldsymbol{x}_k\boldsymbol{x}_k^H]$，在DPC中，$x_1,\cdots,x_K$互不相关，因此$\mathrm{Tr}[\boldsymbol{Q}_x]=\mathrm{Tr}[\boldsymbol{Q}_1+\cdots+\boldsymbol{Q}_K]\leqslant P$。

速率式（14-36）有一个重要特点是，它既不是协方差矩阵的凹函数也不是凸函数。这就很难找出容量域，因为一般要搜索整个满足功率受限条件的协方差矩阵空间[46-47]。不过在14.7节中我们将看到，MIMO广播信道和MIMO多址信道间存在对偶性，可以利用这一点来大大简化容量域的计算。对于$M_t=2$，$M_r=1$的两用户信道，图14-13中画出

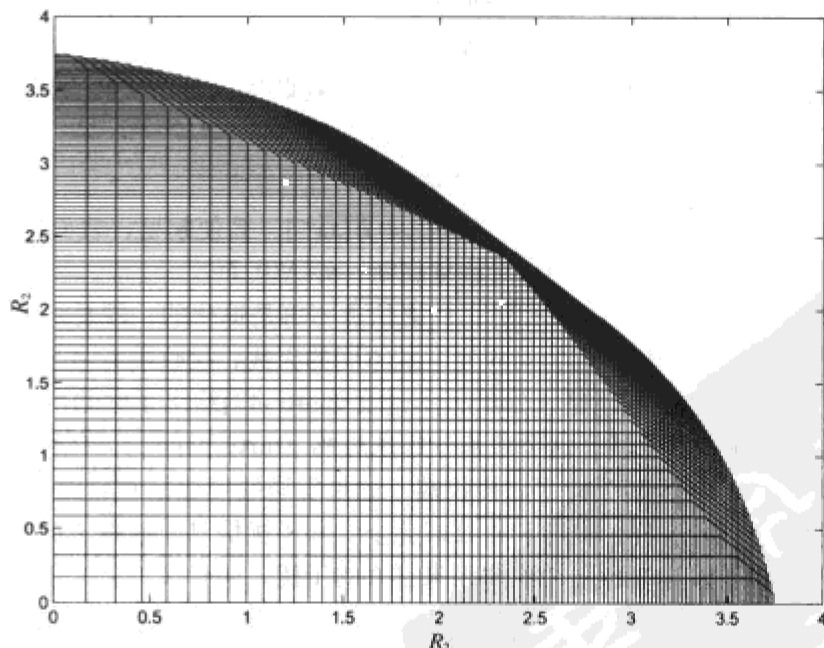


图14-13　MIMO广播信道的容量域，$\boldsymbol{H}_1=[1\ 0.5]$，$\boldsymbol{H}_2=[0.5\ 1]$，$P=10$

了利用对偶特性计算出的容量域。图中轮廓线之内的区域就是容量域，轮廓线内的每条线对应一个与MIMO广播信道对偶的MIMO多址信道的容量域，两者的功率和相同。在14.7节中将看到，这些对偶域的并就是MIMO广播信道的容量域的边界。 484

14.6 上行（多址）信道容量

14.6.1 AWGN信道下的容量

多址信道有K个发送端、一个接收端，每个发送端的功率是P_k，相应的信道增益为g_k。假设这个接收端及所有的发送端都只有一根天线，接收信号叠加了功率谱密度为$N_0/2$的加性高斯白噪声。两用户多址信道的容量域是所有满足下列约束条件（见文献[35]的14.1节）的向量(R_1, R_2)的闭凸包

$$R_k \leqslant B\log_2\left(1+\frac{g_k P_k}{N_0 B}\right) \quad k=1,2 \tag{14-38}$$

$$R_1+R_2 \leqslant B\log_2\left(1+\frac{g_1 P_1+g_2 P_2}{N_0 B}\right) \tag{14-39}$$

第一个约束条件（14-38）是单独每个信道的容量。第二个约束条件（14-39）的意思是说，所有用户的速率和不可能大于这样一个“超级用户”的容量，它的接收功率等于所有用户接收功率的和。对于K个用户，容量域为

$$C_{\text{MAC}}=\left\{(R_1,\cdots,R_K):\sum_{k\in S}R_k \leqslant B\log_2\left(1+\frac{\sum_{k\in S}g_k P_k}{N_0 B}\right)\forall S\subset\{1,2,\cdots,K\}\right\} \tag{14-40}$$

上式说明，对于任意的用户子集，其速率和都不能大于这样一个超级用户的容量，该超级用户的接收功率等于子集中所有用户的接收功率之和。

对于所有容量域内的向量$(R_1,\cdots,R_K)$，$\sum_{k=1}^{K}R_k$的最大值就是多址信道的和速率容量。和广播信道一样，和速率容量是一个比K容量域简明的度量，它反映系统在不考虑公平性时所能达到的最大吞吐量。由式（14-40）可知，让所有用户都用最大功率发送就可以在加性高斯白噪声多址信道中达到和速率容量，即

$$C_{\text{MACSR}}=B\log_2\left(1+\frac{\sum_{k=1}^{K}g_k P_k}{N_0 B}\right) \tag{14-41}$$

从直观来解释就是，多址信道中每个用户的发送功率都是单独约束的，用户不以满功率发送就是浪费。而在AWGN广播信道中，达到和速率容量（14-20）的方法是只发送给信道最好的用户，虽然是一个用户，但因为功率是所有用户共享的，因此并没有浪费功率。

两用户多址信道的容量域示于图14-14中，其中C_1、C_2分别为：

$$C_k=B\log_2\left(1+\frac{g_k P_k}{N_0 B}\right) \quad k=1,2 \tag{14-42}$$

$$C_1^* = B\log_2\left(1+\frac{g_1P_1}{N_0B+g_2P_2}\right) \tag{14-43}$$

485

$$C_2^* = B\log_2\left(1+\frac{g_2P_2}{N_0B+g_1P_1}\right) \tag{14-44}$$

$(C_1, 0)$表示用户1以最大速率发送，用户2不发送时达到的速率向量，反过来就是$(0, C_2)$。折点(C_1, C_2^*)和(C_1^*, C_2)是采用前述叠加码中的串行干扰抵消时可达到的速率向量。用户1以最大速率C_1发送，其信号对用户2来说是噪声，因此用户2能以任意小错误率发送的速率就是C_2^*。接收端从接收信号中减去用户2的信号，剩下的部分就是叠加了噪声的用户1的信号，因此能以任意小的错误率达到速率C_1。这样，(C_1, C_2^*)就是可达到的速率向量，(C_1^*, C_2)也一样。以时分方式采用这两种策略就产生了图中从(C_1, C_2^*)到(C_1^*, C_2)的连线。这个时分线也可以通过*速率分解*（rate splitting）得到。速率分解就是把用户的数据流分成若干子数据流，编码时只把这些子数据流当作来自不同的虚拟用户[51]。注意在广播信道中，信噪比最大的用户总是最后译码，而在MAC中没有这种顺序问题，这是两者的一个基本区别。

若两个用户都用式（14-42）的最大速率按时分方式工作，就得到了连接C_1到C_2的直线。频分的速率与带宽分配有关。设B_1和B_2是分配给两个用户的带宽，可达速率域为

$$\mathcal{C}_{\text{FD}} = \bigcup_{\{B_1,B_2:B_1+B_2=B\}}\left(R_1 = B_1\log_2\left(1+\frac{g_1P_1}{N_0B_1}\right), R_2 = B_2\log_2\left(1+\frac{g_2P_2}{N_0B_2}\right)\right) \tag{14-45}$$

结果画在图14-14中，明显比时分好。实际上，在式（14-45）中取$B_1=\tau B$、$B_2=(1-\tau)B$便有$R_1>\tau C_1$、$R_2>(1-\tau)C_2$。可以证明，频分的曲线在一点上可以达到容量域边界，这个点对应和速率R_1+R_2最大的速率向量（见文献[35]的14.3节）。达到此速率向量时，带宽分配B_1、B_2与相应的接收功率g_1P_1、g_2P_2成正比。

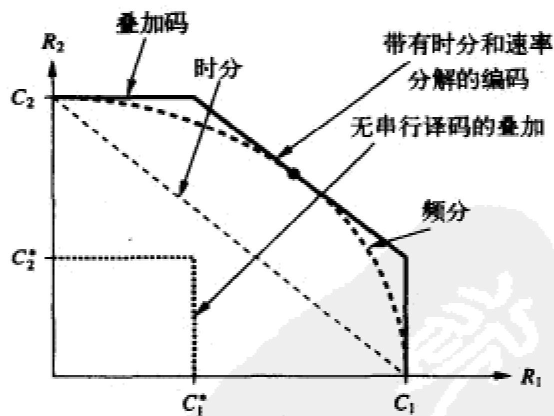


图 14-14

如同广播信道中的时候，通过有效率地使用发送功率，多址信道中TD也能达到FD的速率域。如果每个用户的功率约束P_1、P_2是平均值约束，那么由于用户k使用信道的时间比例只是τ_k，在这段时间内的功率可以提高为P_k/τ_k。这种可变功率TD的速率域为

$$\mathcal{C}_{\text{TD,VP}} = \bigcup_{\{\tau_1,\tau_2:\tau_1+\tau_2=1\}}\left(R_1 = \tau_1B\log_2\left(1+\frac{g_1P_1}{N_0\tau_1B}\right)\ R_2 = \tau_2B\log_2\left(1+\frac{g_2P_2}{N_0\tau_2B}\right)\right) \tag{14-46}$$

486

在式（14-46）中取$B_k \triangleq \tau_k B$就得到与式（14-45）相同的速率域。

也可以不采用串行译码的叠加码，此时每个用户所发送的信号对其他用户来说就是噪声，因此速率域不会超过(C_1^*, C_2^*)，明显在FD和TD的速率域之下。

例14.8：考虑一个两用户的AWGN多址信道，用户的发射功率为$P_1=P_2=100\text{mW}$，信道增益分别为$g_1=0.08$和$g_2=0.001$，系统带宽为$B=100\text{kHz}$，噪声的$N_0=10^{-9}\text{W/Hz}$。求多址信道容量域的折点。若用户2要求的速率分别为$R_2=100\text{kbit/s}$和$R_2=50\text{kbit/s}$，求用户1相应可达到的速率。

解：根据式（14-42）至式（14-44）可得

$$C_1 = B\log_2\left(1+\frac{g_1P_1}{N_0B}\right) = 6.34\times10^5$$

$$C_2 = B\log_2\left(1+\frac{g_2P_2}{N_0B}\right) = 1\times10^5$$

$$C_1^* = B\log_2\left(1+\frac{g_1P_1}{N_0B+g_2P_2}\right) = 5.36\times10^5$$

$$C_2^* = B\log_2\left(1+\frac{g_2P_2}{N_0B+g_1P_1}\right) = 1.77\times10^5$$

用户2的最大速率为100kbit/s，如果他要求的速率为$R_2=100\text{kbit/s}$，那么相应的速率向量是折点(C_1^*, C_2)，此时用户1可达到的速率为$R_1 = C_1^* = 536\,\text{kbit/s}$；如果用户2要求的速率只是50kbit/s，那么速率向量落在容量域的TD部分，即两个折点(C_1, C_2^*)和(C_1^*, C_2)之间的直线$(R_1, R_2)=\tau(C_1, C_2^*)+(1-\tau)(C_1^*, C_2)$，其中$0\leqslant\tau\leqslant1$。能满足$R_2=50\text{kbit/s}$的$\tau$可从$\tau C_2^*+(1-\tau)C_2=R_2$解出为$\tau=(R_2-C_2)/(C_2^*-C_2)=0.51$，大约在两个折点的正中。因此用户1可达到的速率为$R_1=\tau C_1+(1-\tau)C_1^*=5.86\times10^5$。

从本例中可以看出远近效应对多址信道有很大的影响。虽然两个用户有相同的发射功率，由于用户2的信道增益远小于用户1，其速率最大只能到100kbit/s，而用户1的可达速率在536kbp和634kbit/s之间。此外，用户2的信道增益很小，使得它产生的干扰对用户1影响也很小，用户1不受干扰时速率可达到C_1=634kbit/s，受到干扰时速率为$C_1^*=536\text{kbps}$。而用户1对用户2的干扰会严重限制了用户2的速率，使用户2的速率从C_2= 634kbit/s降到了$C_2^*=1.77\text{kbps}$，几乎将低了两个数量级。

487

14.6.2 衰落信道的容量

现在考虑既有衰落，又有加性白高斯噪声的多址信道的容量域，其中每个用户的信道增益都在随时间变化。假设所有用户都只有一个发送天线，接收端也只有一个接收天线，高斯白噪声的功率谱密度为$N_0/2$，每个用户有单独的功率约束$\overline{P}_k, k=1,\cdots,K$，用户$k$在时刻$i$的信道增益为$g_k[i]$，用户间的信道增益相互独立。$i$时刻的信道衰落状态定义为$g[i]=(g_1[i],\cdots,g_k[i])$，在不致混淆的情况下，我们将略去时间$i$。假设接收端和发送端都

有理想的CSI，仅接收端有CSI的情况请参见文献[52]的6.3节。和广播信道、单用户信道一样，多址衰落信道的容量域也分为遍历容量域和中断容量域。遍历容量域是所有衰落状态下平均可达的速率向量；中断容量域是除了某个中断率之外的衰落状态下都可以维持的最大速率向量。

先考虑文献[40]中所导出的遍历容量。令P表示从衰落状态为$g=(g_1,\cdots,g_k)$到各用户的功率分配$P_1(g),\cdots,P_K(g)$之间的映射，它代表一种功率分配策略。令F_{MAC}表示所有满足平均功率约束$\bar{P}_k$的功率分配策略P的集合：

$$\mathcal{F}_{MAC} \triangleq \{\mathcal{P}: E_g[P_k(g)] \leqslant \bar{P}_k, k=1,\ldots,K\}$$

则对于给定的衰落状态g和功率分配$P_1(g),\cdots,P_K(g)$时多址信道的容量域为：

$$\mathcal{C}_{MAC}(P_1(g),\cdots,P_K(g)) = \left\{(R_1,\cdots,R_K): \sum_{k\in S} R_k \leqslant B\log_2\left(1+\frac{\sum_{k\in S} g_k P_k(g)}{N_0 B}\right) \forall S\subset\{1,2,\cdots,K\}\right\} \tag{14-47}$$

在功率分配策略P方案下，所有衰落状态下平均可达的速率集合为：

$$\mathcal{C}_{MAC}(\mathcal{P}) = \left\{(R_1,\cdots,R_K): \sum_{k\in S} R_k \leqslant E_g\left[B\log_2\left(1+\frac{\sum_{k\in S} g_k P_k(g)}{N_0 B}\right)\right] \forall S\subset\{1,2,\cdots,K\}\right\} \tag{14-48}$$

对于所有满足各用户功率约束条件的功率分配策略，其速率域的并集就是遍历容量域：

$$\mathcal{C}_{MAC}(\bar{P}_1,\cdots,\bar{P}_K) = \bigcup_{\mathcal{P}\in\mathcal{F}_{MAC}} \mathcal{C}_{MAC}(\mathcal{P}) \tag{14-49}$$

由式（14-41）、式（14-48）和式（14-49）得到衰落多址信道的和速率容量为：

488

$$C_{MACSR} = \max_{\mathcal{P}\in\mathcal{F}_{MAC}} E_g\left[B\log_2\left(1+\frac{\sum_{k=1}^{K} g_k P_k(g)}{N_0 B}\right)\right] \tag{14-50}$$

用拉格朗日法可求出式（14-50）的最大值。其结果表明，达到和速率容量的方法是在每个衰落状态下只允许加权信道增益g_k/λ_k最大的用户发送[53]，其中λ_k是对应于用户k的平均功率约束条件的拉格朗日系数，它与$\bar{P}_k$及衰落分布有关。根据对称性，如果所有用户的衰落分布及平均功率约束相同，那么拉格朗日系数λ_k也相同，此时最佳的方法就是选择衰落状态g中信道增益g_k最大的用户发送。选择出用户后，功率的分配是沿时间的注水法。这一点的直观解释是：用户可以随时间调整其发送功率，因此让信道条件最好的用户按与其信道条件相适应的功率发送时，系统资源的利用率最高。当用户的平均接收功率不等时，选信噪比最好的用户发送不再是最佳的，因为平均信噪比很小的用户很少有机会发送，这些用户的功率资源未能得到充分利用。

文献[41]给出了多址信道的零中断容量域，定义为所有用户在满足其平均功率约束

的条件下，在所有衰落状态下都能达到的速率集合。根据式（14-40），在功率分配策略$\mathcal{P}$下，衰落状态为g时的多址信道容量域为：

$$\mathcal{C}_{\mathrm{MAC}}(\mathcal{P})=\left\{(R_1,\cdots,R_K):\sum_{k\in S}R_k\leqslant B\log_2\left(1+\frac{\sum_{k\in S}g_kP_k(g)}{N_0B}\right)\ \forall S\subset\{1,2,\cdots,K\}\right\}$$

（14-51）

因此，在功率分配策略为$\mathcal{P}$时，在所有衰落状态g下都可达的速率集合为：

$$\mathcal{C}^0_{\mathrm{MAC}}(\mathcal{P})=\bigcap_g\mathcal{C}_{\mathrm{MAC}}(\mathcal{P})\tag{14-52}$$

对于所有满足用户功率约束的功率分配策略$\mathcal{P}\in\mathcal{F}_{\mathrm{MAC}}$，$\mathcal{C}^0_{\mathrm{MAC}}(\mathcal{P})$的并集就是多址信道的零中断容量域：

$$\mathcal{C}^0_{\mathrm{MAC}}(\bar{P}_1,\cdots,\bar{P}_K)=\bigcup_{\mathcal{P}\in\mathcal{F}_{\mathrm{MAC}}}\bigcap_g\mathcal{C}_{\mathrm{MAC}}(\mathcal{P})\tag{14-53}$$

多址信道的中断容量域与零中断容量域类似，它允许用户以某个中断率在某些衰落状态下停止传输。与广播信道时的情形一样，多址信道的中断容量域要比零中断容量域更难求出，因为在任意给定的衰落状态下，发送端必须要决定出哪个用户被中断、非中断用户的译码顺序以及非中断用户的发送功率。文献[43]间接得到了能满足用户功率约束和用户中断率约束的中断容量域。对于多址信道，也可以把遍历容量和中断容量结合而得到最小速率容量域。这个最小速率容量域也和广播信道的情形一样，它是所有衰落状态下平均的可达速率向量，且每个向量在所有衰落状态中能以某个中断率（包括零中断率）大于一定的最小值。文献[54]利用广播信道和多址信道的对偶原理得出了多址信道的最小速率容量域。对偶原理将在下节介绍。

489

14.6.3 多天线时的容量

现在考虑多天线的多址信道。如同MIMO广播信道时，我们把带宽归一化为$B=1\mathrm{Hz}$，并假设接收端的噪声向量$\boldsymbol{n}$为循环对称复高斯，$\boldsymbol{n}: N(0,\boldsymbol{I})$。令$M_r\times M_t$维矩阵$\boldsymbol{H}_k$表示用户$k$到接收端的MIMO信道增益。定义$\boldsymbol{H}=[\boldsymbol{H}_1,\cdots,\boldsymbol{H}_K]$，则用户$k$的功率约束为$P_k$时的高斯MIMO多址信道的容量域为[55-57]：

$$\begin{aligned}&\mathcal{C}_{\mathrm{MAC}}((P_1,\cdots,P_K);\boldsymbol{H})\\&=\bigcup_{\{\boldsymbol{Q}_k\geq 0,\ \mathrm{Tr}(\boldsymbol{Q}_k)\leq P_k\,\forall k\}}\Bigg\{(R_1,\cdots,R_K):\\&\qquad\sum_{k\in S}R_k\leqslant\log_2\det\left[\boldsymbol{I}+\sum_{k\in S}\boldsymbol{H}_k\boldsymbol{Q}_k\boldsymbol{H}_k^H\right]\forall S\subseteq\{1,\cdots,K\}\Bigg\}\end{aligned}\tag{14-54}$$

得到这个容量域的方法是：用户k发送一个零均值高斯随机向量$\boldsymbol{x}_k$，其空间协方差矩阵为$\boldsymbol{Q}_k=E\left[\boldsymbol{x}_k\boldsymbol{x}_k^H\right]$。每一组协方差矩阵$(\boldsymbol{Q}_1,\cdots,\boldsymbol{Q}_K)$对应一个$K$维的多面体（即$\{(R_1,\cdots,R_K)$:

$\sum_{k\in S}R_k \leqslant \log_2\det[I+\sum_{k\in S}H_kQ_kH_k^H]\forall S\subseteq\{1,\cdots,K\}\}$)。对于所有满足功率约束的协方差矩阵，容量域就是这些多面体的并集。通过串行译码陆续减去已经检测出的信号便可以达到容量域的角点。与单天线相比，式（14-54）也是基于用户子集的速率和定义的，也是通过串行译码达到容量域的角点。

在两用户系统中，每一组协方差矩阵对应一个多边形，形式上与单天线多址信道的容量域类似。例如对于 $R_1=\log_2\det\left[I+H_1Q_1H_1^H\right]$，

$$R_2=\log_2\det[I+H_1Q_1H_1^H+H_2Q_2H_2^H]-R_1$$
$$=\log_2\det[I+(I+H_1Q_1H_1^H)^{-1}H_2Q_2H_2^H]$$

490 所对应的角点，用户2先译码（用户1为干扰），用户1后译码（用户2干扰被消除）。

14.7 上下行信道的对偶性

从图14-1来看，上行信道和下行信道几乎只是箭头的方向不同。不过，它们之间有三个本质的区别：首先，下行信道中每个接收端都有一个加性噪声项，而上行信道中只有一个噪声项（只有一个接收机）；其次，下行信道只有一个发送功率约束，而上行信道中每个用户都有各自的功率约束；最后，下行信道中每个用户接收到的信号和干扰都经过相同的信道，而上行信道中这些信号经过了不同的信道，因此有远近效应。尽管对单独的上行信道和下行信道已经有了大量的研究，但对它们的联系，以及利用这种联系来进行分析和设计的研究还很少。本节讨论这两个信道的对偶性，并说明这种对偶性在容量分析及上下行发送设计中的应用。

称图14-1所示的K用户上行信道和下行信道是互为对偶的，若其满足以下三个条件：

(1) 对所有用户，上下行的信道冲激响应$h_k(t), k=1,\cdots,K$ 相同。对于AWGN信道或平衰落信道，此即双向的信道增益相同。

(2) 下行信道中每个信道噪声的统计特性都与上行信道相同。

(3) 下行信道的功率约束P等于上行信道中各用户的功率约束$P_k(k=1,\cdots,K)$之和。

虽然上行（多址）信道和下行（广播）信道有这样的相似性，它们的容量域却相当不同。例如图14-10所示的两用户的AWGN广播信道的容量域（图中的最大速率域）与图14-14所示的两用户AWGN多址信道的容量域明显不一样。对于衰落信道，无论是遍历容量、中断容量还是最小和容量，两者也都有很大的差别。不过，容量域的形状虽然不同，但达到容量域的方法都是叠加码，最佳译码也都是串行译码和干扰抵消。

基于编译码的这种相似性，通过如下两点就能建立起二者的对偶关系：(i) 使广播信道中的功率约束等于多址信道中用户功率约束之和；(ii) 调整广播信道的增益使它产生多址信道中的远近效应。文献[54]证明：若已知广播信道或多址信道中某一个容量域和最佳发送策略，另一个的容量域和最佳发送策略可以通过对偶关系得到。特别对于AWGN信道，文献[54]证明，功率约束为P的广播信道和功率约束满足 $\sum_{k=1}^{K}P_k\leqslant P$的对偶多址信道容量域相同。多址信道这种和功率约束表示各用户的发射功率从一个总量为

P的公共功率池中分配。广播信道的容量域可以表示为所有满足和功率约束的对偶多址信道的并集[54]：

$$\mathcal{C}_{BC}(P,g)=\bigcup_{\{(P_1,\cdots,P_K):\sum_{k=1}^{K}P_k=P\}}\mathcal{C}_{MAC}(P_1,\cdots,P_K;g) \tag{14-55}$$

式中$\mathcal{C}_{BC}(P,g)$是总功率约束为P，信道增益为$g=(g_1,\cdots,g_k)$，噪声为$n_k=N_0/g_k$时AWGN广播信道的容量域，即式（14-18）。$\mathcal{C}_{MAC}(P_1,\cdots,P_K;g)$是信道增益为$g=(g_1,\cdots,g_k)$，各用户的功率限制为$P_1,\cdots,P_k$的AWGM多址信道的容量域，即式（14-40）。这种关系如图14-15所示，图中满足和功率约束的对偶多址信道的容量域的并集形成了广播信道的容量域。 491

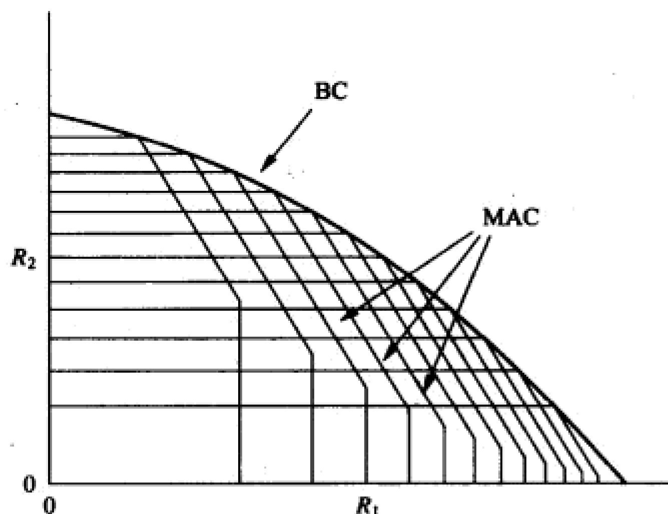


图14-15 AWGN广播（下行）信道的容量域是对偶多址（上行）信道容量域的并集

除了式（14-55）这种关系外，还可以证明[54]，广播信道容量域的边界点对应的最佳功率分配也可以从与其相交的对偶多址信道的功率分配得到。并且在这个交点处，广播信道的译码顺序与对偶多址信道相反。因此，广播信道的最佳编码和译码都可以从对偶多址信道中得到，这一点对实际设计有启发意义。

我们也能从对偶的广播信道得到多址信道的容量域，这一点是通过调整信道增益实现的。从式（14-40）可见，将用户k的信道增益g_k乘上一个权值α，同时给它的发送功率乘上$1/\alpha$，则容量域不变。但这样一来，对偶的广播信道中叠加码的编译码顺序将因信道增益的改变而改变，其容量域也因此而不同。文献[54]证明，所有因子α_k下，对偶广播信道容量域的交集就是多址信道的容量域，即：

$$\mathcal{C}_{MAC}(P_1,\cdots,P_K;g)=\bigcap_{(\alpha_1,\cdots,\alpha_K)>0}\mathcal{C}_{BC}\left(P=\sum_{k=1}^{K}\frac{P_k}{\alpha_k};g=(\alpha_1 g_1,\cdots,\alpha_K g_K)\right) \tag{14-56}$$

492

式中$\mathcal{C}_{MAC}(P_1,\cdots,P_K;g)$和$\mathcal{C}_{BC}(P,g)$与式（14-55）中的定义相同。这种关系示于图14-16中，图中对第一个用户乘以不同的因子[1]后得到的不同广播信道的容量域的交集形成了

1. 求交集时只需对$K-1$个用户进行加权，因子$(\alpha_1,\cdots,\alpha_{K-1},\alpha_K)$等效于$(\alpha_1/\alpha_K,\cdots,\alpha_{K-1}/\alpha_K,1)$。

多址信道的容量域。当$\alpha \to 0$时，用户1的信道增益αg_1趋近于零，但总功率$P = P_1/\alpha + P_2$也趋向于无穷大，由于用户2的信道增益没变，使其容量也随总功率而渐进增大；而当$\alpha \to \infty$时情况正相反，用户1的信道增益增大，总功率$P = P_1/\alpha + P_2 \geqslant P_2$，使用户1的容量随信道增益的增长渐进增大。因子$\alpha$从零到无穷之间变化得到不同的广播信道容量域，它们相交形成多址信道的容量域。$\alpha = g_2/g_1$时的广播信道下两用户的信道增益相同，得到多址信道容量域的时分线。对于多址信道容量域上的某个点，其最佳译码顺序也可以从相交于该点的对偶广播信道得到。

493

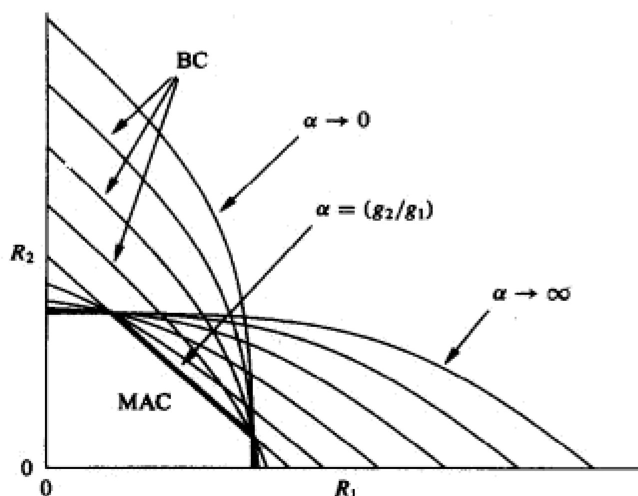


图 14-16

文献[54]还将对偶性扩展到其他一些重要的信道，特别是平衰落的多址信道和广播信道，使得一个信道的遍历容量域、中断容量域、最小速率容量域以及相应的编译码都可以从其对偶信道得到。在并行或频率选择性的多址和广播信道，容量域的对偶性也成立[58-59]。对偶性还可以应用到多天线（MIMO）的多址信道和广播信道中[48],[60-61]。文献[60]将对偶的概念推广到MIMO系统，证明功率约束为P、信道增益矩阵为$\boldsymbol{H}$的MIMO广播信道的容量域等于信道增益矩阵为$\boldsymbol{H}^H$、功率和约束为P的对偶多址信道容量域的并集，即：

$$\mathcal{C}_{\mathrm{BC}}(P,\boldsymbol{H}) = \bigcup_{(P_1,\ldots,P_K):\sum_{k=1}^{K}P_k=P} \mathcal{C}_{\mathrm{MAC}}((P_1,\ldots,P_K);\boldsymbol{H}^H)$$

其中$\mathcal{C}_{\mathrm{BC}}(P,\boldsymbol{H})$和$\mathcal{C}_{\mathrm{MAC}}(P_1,\cdots,P_K;\boldsymbol{H}^H)$分别由式（14-37）和式（14-54）给出。对偶关系示于图14-13，其中轮廓线内的区域是MIMO广播信道的容量域，轮廓线之内所画的各个域是在功率和同为P的不同功率分配的MIMO多址信道的容量域。前面已经提到，MIMO广播信道的容量域很难求解，因为对于待优化的协方差来说，容量域既不是凸函数也不是凹函数。而MIMO多址信道中的容量域是一个标准的凸优化问题[62]。除了容量域关系外，对偶性也把多址信道和广播信道的最佳发送方案联系到一起。对于MIMO广播信道容量域内的一个点，相应有一个能达到这个点的对偶变换的多址信道，于是该广

播信道的最佳发送方案可从这个对偶多址信道得到。因此，对偶性可以简化MIMO信道的容量及最佳发送方案的求解。

14.8 多用户分集

多用户分集利用了这样一个事实，即在一个多用户系统中，若各用户的信道衰落相互独立，则任何时刻都会有一部分用户的信道条件好于其他用户。在任意给定的时刻，只向信道条件最好的那些用户发送，就是把系统资源就分配给了最有效率的用户，从而能提高系统的容量或性能。文献[53]最早提出在上行信道中用多用户分集来提高吞吐量、减小误码率。这一思想同样也可以用在下行信道中。多用户分集是第7章所介绍的单用户分集的扩展。在单用户分集系统中，一个点到点的链路由几个独立信道组成，合并这些信道上的信号就可以提高系统的性能。在多用户分集中，多个信道对应着不同的用户，因此系统一般采用选择分集选出信道条件最好的用户。多用户分集增益来自于各用户信道条件的差异，衰落的动态范围越大，多用户分集增益也越大。此外，多用户分集也和其他分集一样，性能的改善程度随独立信道数的增加而增加，它在有大量用户的系统中最为有效。 494

14.5节已经表明，能最大化下行信道吞吐量（和速率容量）的方法是把系统的全部带宽分配给信道条件最好的用户。对于上行信道，如果各用户有相同的衰落分布和相同的平均功率，14.6节中也有同样的结论；如果用户的衰落分布不同或平均功率不同，则是选择加权信道增益最大的用户，加权值取决于信道的增益、衰落分布和平均功率约束条件。根据用户的信道条件来调度用户发送的方法称为*机会调度*（opportunistic scheduling）。文献[34]和文献[53]中的数值结果表明，结合使用功率控制时，机会调度能显著增加以和速率容量衡量的上行和下行的吞吐量。

机会调度也能改善误码性能[53]。令$\gamma_k[i]$, $k=1,\cdots,K$表示i时刻每个用户的信噪比。只向最大信噪比的用户发送时，系统在i时刻的信噪比为$\gamma=\max_k\gamma_k[i]$。当K充分大时，独立同分布瑞利信道下，这个信噪比大约是单用户信噪比的$\ln K$倍[63]，即多用户分集大致有$\ln K$倍的信噪比增益。另外，系统的误码率$P_s(\gamma)$也如单用户系统中的选择分集（见7.2.2节）一样有分集增益，当用户数逐步增大时，误码率也逐渐接近于AWGN信道的误码率。

在无线系统中选信道条件最好的用户进行调度发送存在两个问题：公平性和延时。如果信道衰落非常慢的话，某一个用户可能会长期占用信道，使其他用户相邻两次发送的时间间隔相当长，这样的时延对于某些应用是不能接受的。此外，平均信噪比很低的用户的信道变成瞬时信噪比最高的机会也很低，发送的机会也很少，造成了系统资源分配的不公平。为此，文献[63]提出了一种称为*正比公平调度*（proportional fair scheduling）的解决方法。在i时刻，若假设把所有带宽和功率分配给用户k时，它可支持的最大下行速率为$R_k[i]$。令$T_k[i]$为用户k在i时刻实际实现的指数加权平均吞吐量，平均滤波器的参数是i_c。那么正比公平调度选择$R_k[i]/T_k[i]$最大的用户在i时刻发送。如果i时刻所有用户的吞吐量都相同，即$T_k[i]=T[i]$，系统调度到的就是信道条件最好的用户。如果某个用户（比如用户j）的吞吐量远小于其他用户的吞吐量，即$T_j[i]=T_k[i]$, $j\neq k$，则

$R_j[i]/T_j[i]$可能很大，系统就很可能调度到用户j。假设i时刻用户k^*的$R_k[i]/T_k[i]$值最大，则下一时隙的吞吐量更新为：

$$T_k[i+1]=\begin{cases}(1-1/i_c)T_k[i]+(1/i_c)R_k[i] & k=k^*\\(1-1/i_c)T_k[i] & k\neq k^*\end{cases}\tag{14-57}$$

这种调度方式在用户的吞吐量相对公平时选择信道条件最好的用户，而当某个用户的吞吐量很差时将优先调度这个用户，直到其吞吐量能达到一种相对的平衡。正比公平调度的吞吐量明显要低于总选择最好用户的方法（其吞吐量最大），用户的平均信道质量相差越大时吞吐量的损失也越多。正比公平调度的时延可通过参数i_c来控制，其时延随i_c的增大而增大，同时吞吐量也随i_c增大。若各用户的平均信噪比相同，则i_c趋于无穷时的正比公平调度简化为调度信道最好的用户。CDMA 2000系统中的分组数据采用了正比公平调度[64]，文献[65]给出了相应的性能评估。文献[66~67]研究了综合考虑公平性和时延限制的机会调度问题，并给出了存在实际限制（如信道估计不理想）时的性能。

495

14.9 MIMO多用户系统

配置有多个收发天线的多用户系统称为MIMO多用户系统。多个天线能显著提高系统性能。利用多天线的方法有多种，可以用作分集以提高误码率性能、可以通过复用来提高容量域、还可以通过空间方向性来减少用户间的干扰。一般需要综合考虑这三点[68]。

MIMO多用户系统的复用增益反映增加天线所带来的上行或下行容量域的增加。MIMO对单用户容量的显著提高促使人们对MIMO多用户信道的容量域开展了大量的研究。14.6.3节和14.5.5节已经介绍了AWGN MIMO系统的上行和下行容量域，将这些结果加以扩展就可以得到收发都有理想CSI的MIMO多用户衰落信道的容量域。文献[69]给出了容量结果，并给出了不同CSI假设下一些尚未解决的问题。

在10.4节中我们已经介绍了单用户系统中的波束成形技术，它以一定的容量损失为代价，达到了满分集。在多用户系统中，多用户分集的作用使波束成形的容量损失较小，特别在用户数非常多时，波束成形能达到MIMO下行信道的和速率容量[70~71]。

14.8节讨论了单天线时的多用户分集技术，它能利用不同用户信道质量之间的差异来提高系统性能。多用户分集用于多天线系统时有两方面的好处：首先，只有信道条件最好的那些用户在传输，所以MIMO多用户分集提高了信道质量；其次，MIMO多用户分集系统可以产生许多空间方向来对准所选择的用户，所以在一个给定状态下资源分配对用户进行选择时不仅是要选择信道质量好的，而且要选择空间分割好的，从而可以减小用户间干扰。这两点使得用户数较大时，采用相对简单的次优技术也能接近于最优性能[70][72]。它也使上行只用单个发送天线、下行只用单个接收天线就能达到增加容量的目的，从而可以降低移动终端的复杂性和功耗。实际上，MIMO广播信道的和速率容量与发送天线数和所有用户接收天线总数两者之间的较小值大致成线性增长[73]，给定基站发送天线数并且用户数量很大时，终端增加天线并不能提高系统的容量。MIMO多址信道也一样，其和速率容量增益也随接收天线数和各用户发送天线总数中的较小值大致成线性增长，在基站有大量接收天线且系统用户数很多时，增加用户端的天线数同样无助于

容量的提高。多用户分集增益随衰落的动态范围和衰落速率的增加而增加，有意识地对多个发送天线进行幅度和相位调制可以增加信道的动态范围和衰落速率，从而能得到更高的多用户分集增益，这种技术称为*机会波束成形*（opptunistic beamforming）[63]。

496

人们还提出了用于MIMO多用户系统的空时编码和空时调制技术[74-76]。这些技术能充分实现分集、复用和方向性的各种折中考虑。多用户检测技术也可扩展到MIMO系统，它能使性能显著提高[77-79]。在宽带信道中，多用户MIMO需要解决频率选择性衰落的问题[76][80-81]。由于宽带系统中的频率选择性衰落是一种形式的分集，采用一些先进的技术后能获得比窄带信道更显著的性能增益。不过在低信噪比时，多天线的主要作用是收集能量。MIMO多用户系统面临的主要挑战是以合理的复杂度设计出能在各种实际环境中达到理论性能增益的信号传输技术。

习题

14.1 某FDMA系统同时向100个用户发送多媒体数据，每个用户需要的带宽为10MHz。所分配频带的两侧需有1MHz的保护带宽以减少带外干扰。问此系统需要的总带宽是多少？

14.2 GSM系统上下行各分配有25MHz的频带，分为125个TDMA载波，每载波分为8个时隙。这8个时隙构成一个GSM帧。每帧有一些前导和后置比特。每个时隙按时间次序包括3个起始比特、58个数据比特、26个均衡器训练比特、58个数据比特、3个终止比特，最后是8.25个比特时间的保护间隔。数据传输速率为270.833kbit/s。

(a) 画出GSM的帧及时隙结构。

497

(b) 求出一个时隙中数据所占的比例以及用户的信息数据速率。

(c) 不考虑帧的前导和后置比特，求一帧的时间及同一用户前后两个时隙间的延迟。

(d) 信道的最大时延扩展为多少时，保护间隔及终止比特能防止时隙重叠？

14.3 某干扰受限的DS-CDMA系统的带宽是10MHz，扩频增益$G = 100$，扩频码的互相关系数为$1/G$。

(a) 对于多址信道，假设所有用户具有理想功率控制，使得它们到达接收端的功率相同，求以G和用户数K为参数的接收信号SIR表达式。

(b) 假设采用BPSK调制，目标误比特率为10^{-3}，并假设可将干扰视为加性白高斯噪声。请利用（a）中的SIR表达式求出系统可支持的最大用户数K，并与系统带宽和信息带宽均相同的FDMA作比较。

(c) 话音激活因子表示用户通话时间的百分比，考虑话音激活时，干扰功率应乘以激活因子。请修正（a）中SIR表达式以体现话音激活的影响。并求出能使CDMA系统与FDMA系统用户容量相同的激活因子。结果得到的激活因子数值是否合理？

14.4 设有一个采用FSK调制的FH-CDMA系统，其信息带宽和扩频带宽与14-3题相同，即系统有$G = 100$个频隙，每个频隙的带宽为100kHz。跳频码随机均匀分布，因此某一用户跳至任一频隙的概率均为0.01。和题14-3一样忽略噪声，则某个频隙上只有一个用户时错误率为零。还假设有理想功控，使得接收端每个用户的功率相同。

(a) 若有m个用户同时跳到同一频隙，求误比特率表达式。

(b) 假设系统有K个用户，求多个用户跳到同一频隙的概率。

(c) 求以用户数K为参数的平均误比特率表达式。

14.5　计算纯ALOHA和时隙ALOHA的最大吞吐量T及其对应的负载L。

14.6　某个纯ALOHA系统的传输速率为R = 10Mbit/s，分组长度为1000比特，泊松到达速率是$\lambda = 10^3$。计算负载L、系统吞吐量T以及有效数据速率（成功接收的比特率）。还有哪个L值也能产生完全相同的吞吐量？

14.7　某个三用户上行信道中，各用户的信道增益分别为$g_1 = 1$、$g_2 = 3$、$g_3 = 5$，接收端的噪声为$n = 1$，各用户要求的SINR均为10dB。

(a) 验证式（14-6）的向量方程$(\boldsymbol{I}-\boldsymbol{F})\boldsymbol{P} \geqslant \boldsymbol{u}$等价于各用户的SINR约束条件。

(b) 假设$\rho = 0.1$，请问是否存在一个功率向量能使所有的用户都满足SINR要求？若是，498 则求出发送功率最小并能满足SINR要求的最佳的功率向量P^*。

14.8　假设两用户广播信道的发射功率为P = 10mW、B = 100kHz、$N_0 = 10^{-9}$W/Hz，求采用叠加码时的容量域。

14.9　证明在AWGN广播信道下，将所有功率都分给信道增益最大的用户能达到和速率容量。

14.10　对于衰落广播信道，请推导出能使和速率最大的最佳功率分配式。

14.11　假设两用户衰落广播信道中用户的衰落相互独立，并假设（i）每个用户的接收功率为10mW，（ii）每个用户的有效噪声功率各以0.5的概率取值1mW或5mW。求和速率容量。

14.12　假设两用户衰落广播信道中，B=100kHz，$N_0=10^{-9}$W/Hz，每个用户都经历瑞利衰落，平均接收功率均为P=10mW。求和速率容量。

14.13　设某频分的衰落广播信道在给定功率分配$\mathcal{P}$和带宽分配$\mathcal{B}$时的速率域为式(14.30)所给出的$\mathcal{C}_{FD}(\mathcal{P}, \mathcal{B})$。任意给定一个$\mathcal{C}_{FD}(\mathcal{P}, \mathcal{B})$中的速率向量，求时分情况下能达到相同速率向量的时隙分配和功率分配。

14.14　某时变广播信道的总带宽为B=100kHz，总发送功率为P=10W。用户1的有效噪声以概率3/4为$n_1 = 10^{-5}$W/Hz，以概率1/4为$n_1 = 2\times10^{-5}$W/Hz，用户2的有效噪声以对半的概率为$n_2 = 10^{-5}$W/Hz或$n_2 = 2\times10^{-5}$W/Hz。n_1和n_2相互独立。

(a) 写出n_1和n_2取值的各种可能组合及其概率。

(b) 求时分方式下两用户间的最佳功率分配及其相应的时变容量速率域，假设：(i) 给用户k分配一个在所有时间都固定的时隙比例$\tau_k(\tau_1+\tau_2=1)$以及按时间平均固定的发送功率P；(ii) 只要保持平均功率为P，用户可以在不同时隙内用不同的功率发送。假设发送功率不是平分，试找出一个能超出这个容量域的速率向量。

(c) 再假设频分，设系统带宽以固定的方式在两个用户之间平分，即$B_1 = B_2 = B/2$。再假设平均功率也平分$(P_1 = P_2 = P/2)$，但只要平均功率不变，瞬时功率可以改变。请问可达的最佳速率向量是什么？假如带宽或者功率并不是平分，试找出一个超出这个容量域的速率向量。

499 (d) 速率向量$(R_1$=100000, R_2=100000)是否在该信道的零中断容量域内？

14.15　对于K用户的AWGN多址信道，证明：若将用户k信道增益g_k乘以α，发送功率P_k乘以$1/\alpha_k$，则容量域不变。

14.16 某个两用户多址信道中，系统总带宽为B = 100kHz，接收机噪声的功率谱密度为0.001μW/Hz。用户1的发射功率为P_1 = 3mW，用户2的发射功率为P_2 = 1mW。假设可以忽略路径损耗、衰落和阴影效应。

（a）假设用户1需要300kbps的速率来传送视频，请问时分方式下用户2能分到的速率最大是多少？如果采用有串行干扰抵消的叠加码呢？

（b）求频分速率域与有干扰消除的码分(G = 1)速率域的交点(R_1, R_2)。

（c）分别对时分、有干扰消除的扩频码分、无干扰消除的扩频码分，求满足R_1 = R_2的速率对(R_1, R_2)，其中扩频增益为G=10。（注：对于G>1的容量域，需要用广播信道中相关的容量域推导。）

14.17 证明：达到AWGN多址信道的和速率容量的方法是让所有用户都满功率发送。

14.18 推导出两用户衰落多址信道中，达到和速率向量的最佳功率分配。

14.19 求独立衰落的两用户多址信道的和速率容量。假设（i）每个用户的接收功率为10mW；（ii）每个用户的有效噪声功率以等概率取值于1mW或5mW。

14.20 某三用户下行信道的系统带宽为100kHz，三个用户的衰落在每个时隙都是独立同分布的。如果将全部带宽和功率都分给一个用户，该用户的接收信噪比的可能值为5dB、10dB和20dB，概率均为1/3。

（a）若在每个时隙都将全部带宽和功率分配给信道条件最好的用户，求最大吞吐量。

（b）仿真i_c = 1, 5, 10时正比公平调度算法的吞吐量。

参考文献

[1] D. Bertsekas and R. Gallager, *Data Networks*, 2nd ed., Prentice-Hall, Englewood Cliffs, NJ, 1992.

[2] T. S. Rappaport, *Wireless Communications – Principles and Practice*, 2nd ed., Prentice-Hall, Englewood Cliffs, NJ, 2001.

[3] G. Leus, S. Zhou, and G. B. Giannakis, "Orthogonal multiple access over time- and frequency-selective channels," *IEEE Trans. Inform. Theory*, pp. 1942–50, August 2003.

500

[4] S. Verdú, "Demodulation in the presence of multiuser interference: Progress and misconceptions," in D. Docampo, A. Figueiras, and F. Perez-Gonzalez, Eds., *Intelligent Methods in Signal Processing and Communications*, pp. 15–46, Birkhäuser, Boston, 1997.

[5] W. Stallings, *Wireless Communications and Networks*, 2nd ed., Prentice-Hall, Englewood Cliffs, NJ, 2005.

[6] B. Gundmundson, J. Sköld, and J. K. Ugland, "A comparison of CDMA and TDMA systems," *Proc. IEEE Veh. Tech. Conf.*, pp. 732–5, May 1992.

[7] M. Gudmundson, "Generalized frequency hopping in mobile radio systems," *Proc. IEEE Veh. Tech. Conf.*, pp. 788–91, May 1993.

[8] K. S. Gilhousen, I. M. Jacobs, R. Padovani, A. J. Viterbi, L. A. Weaver, Jr., and C. E. Wheatley III, "On the capacity of a cellular CDMA system," *IEEE Trans. Veh. Tech.*, pp. 303–12, May 1991.

[9] P. Jung, P. W. Baier, and A. Steil, "Advantages of CDMA and spread spectrum techniques over FDMA and TDMA in cellular mobile radio applications," *IEEE Trans. Veh. Tech.*, pp. 357–64, August 1993.

[10] J. Chuang and N. Sollenberger, "Beyond 3G: Wideband wireless data access based on OFDM

and dynamic packet assignment," *IEEE Commun. Mag.*, pp. 78–87, July 2000.

[11] K. R. Santhi, V. K. Srivastava, G. SenthilKumaran, and A. Butare, "Goals of true broadband's wireless next wave (4G-5G)," *Proc. IEEE Veh. Tech. Conf.*, pp. 2317–21, October 2003.

[12] M. Frodigh, S. Parkvall, C. Roobol, P. Johansson, and P. Larsson, "Future-generation wireless networks," *IEEE Wireless Commun. Mag.*, pp. 10–17, October 2001.

[13] E. Anderlind and J. Zander, "A traffic model for non-real-time data users in a wireless radio network," *IEEE Commun. Lett.*, pp. 37–9, March 1997.

[14] K. Pahlavan and P. Krishnamurthy, *Principles of Wireless Networks: A Unified Approach*, Prentice-Hall, Englewood Cliffs, NJ, 2002.

[15] N. Abramson, "The ALOHA system – Another alternative for computer communications," *Proc. Amer. Federation Inform. Proc. Soc. Fall Joint Comput. Conf.*, pp. 281–5, November 1970.

[16] A. Chockalingam and M. Zorzi, "Energy consumption performance of a class of access protocols for mobile data networks," *Proc. IEEE Veh. Tech. Conf.*, pp. 820–4, May 1998.

[17] S. Haykin and M. Moher, *Modern Wireless Communications*, Prentice-Hall, Englewood Cliffs, NJ, 2005.

[18] P. Karn, "MACA: A new channel access method for packet radio," *Proc. Comput. Network Conf.*, pp. 134–40, September 1990.

[19] V. Bharghavan, A. Demers, S. Shenkar, and L. Zhang, "MACAW: A media access protocol for wireless LANs," *Proc. ACM SIGCOMM*, vol. 1, pp. 212–25, August 1994.

[20] *IEEE Standard for Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*, IEEE Standard 802.11, 1997.

[21] Z. J. Haas, J. Deng, and S. Tabrizi, "Collision-free medium access control scheme for ad hoc networks, *Proc. Military Commun. Conf.*, pp. 276–80, 1999.

[22] S.-L. Wu, Y.-C. Tseng, and J.-P. Sheu, "Intelligent medium access for mobile ad hoc networks with busy tones and power control," *IEEE J. Sel. Areas Commun.*, pp. 1647–57, September 2000.

[23] N. Abramson, "Wide-band random-access for the last mile," *IEEE Pers. Commun. Mag.*, pp. 29–33, December 1996.

[24] D. J. Goodman, R. A. Valenzuela, K. T. Gayliard, and B. Ramamurthi, "Packet reservation multiple access for local wireless communications," *IEEE Trans. Commun.*, pp. 885–90, August 1989.

[25] N. B. Mehta and A. J. Goldsmith, "Effect of fixed and interference-induced packet error probability on PRMA," *Proc. IEEE Internat. Conf. Commun.*, pp. 362–6, June 2000.

[26] P. Agrawal, "Energy efficient protocols for wireless systems," *Proc. Internat. Sympos. Pers., Indoor, Mobile Radio Commun.*, pp. 564–9, September 1998.

[27] J. Zander, "Performance of optimum transmitter power control in cellular radio systems," *IEEE Trans. Veh. Tech.*, pp. 57–62, February 1992.

501

[28] G. J. Foschini and Z. Miljanic, "A simple distributed autonomous power control algorithm and its convergence," *IEEE Trans. Veh. Tech.*, pp. 641–6, November 1993.

[29] S. A. Grandhi, R. Vijayan, and D. J. Goodman, "Distributed power control in cellular radio systems," *IEEE Trans. Commun.*, pp. 226–8, February–April 1994.

[30] E. Seneta, *Nonnegative Matrices and Markov Chains*, Springer-Verlag, New York, 1981.

[31] N. Bambos, S. C. Chen, and G. J. Pottie, "Channel access algorithms with active link protection for wireless communication networks with power control," *IEEE/ACM Trans. Network.*, pp. 583–97, October 2000.

[32] S. Kandukuri and N. Bambos, "Power controlled multiple access (PCMA) in wireless communication networks," *Proc. IEEE Infocom Conf.*, pp. 386–95, March 2000.

[33] P. P. Bergmans and T. M. Cover, "Cooperative broadcasting," *IEEE Trans. Inform. Theory*, pp. 317–24, May 1974.

[34] L. Li and A. J. Goldsmith, "Capacity and optimal resource allocation for fading broadcast chan-

nels – Part I: Ergodic capacity," *IEEE Trans. Inform. Theory*, pp. 1083–1102, March 2001.
[35] T. Cover and J. Thomas, *Elements of Information Theory*, Wiley, New York, 1991.
[36] L.-F. Wei, "Coded modulation with unequal error protection," *IEEE Trans. Commun.*, pp. 1439–49, October 1993.
[37] P. P. Bergmans, "A simple converse for broadcast channels with additive white Gaussian noise," *IEEE Trans. Inform. Theory*, pp. 279–80, March 1974.
[38] R. Pickholtz, L. Milstein, and D. Schilling, "Spread spectrum for mobile communications," *IEEE Trans. Veh. Tech.*, pp. 313–22, May 1991.
[39] S. Verdú, *Multiuser Detection*, Cambridge University Press, 1998.
[40] D. Tse and S. Hanly, "Multiaccess fading channels – Part I: Polymatroid structure, optimal resource allocation and throughput capacities," *IEEE Trans. Inform. Theory*, pp. 2796–2815, November 1998.
[41] S. Hanly and D. Tse, "Multiaccess fading channels – Part II: Delay-limited capacities," *IEEE Trans. Inform. Theory*, pp. 2816–31, November 1998.
[42] L. Li and A. J. Goldsmith, "Capacity and optimal resource allocation for fading broadcast channels – Part II: Outage capacity," *IEEE Trans. Inform. Theory*, pp. 1103–27, March 2001.
[43] L. Li, N. Jindal, and A. J. Goldsmith, "Outage capacities and optimal power allocation for fading multiple access channels," *IEEE Trans. Inform. Theory*, pp. 1326–47, April 2005.
[44] N. Jindal and A. J. Goldsmith, "Capacity and optimal power allocation for fading broadcast channels with minimum rates," *IEEE Trans. Inform. Theory*, pp. 2895–2909, November 2003.
[45] E. Larsson and P. Stoica, *Space-Time Block Coding for Wireless Communications*, Cambridge University Press, 2003.
[46] G. Caire and S. Shamai, "On the achievable throughput of a multiantenna Gaussian broadcast channel," *IEEE Trans. Inform. Theory*, pp. 1691–1706, July 2003.
[47] W. Yu and J. M. Cioffi, "Trellis precoding for the broadcast channel," *Proc. IEEE Globecom Conf.*, pp. 1344–8, November 2001.
[48] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the Gaussian MIMO broadcast channel," *Proc. IEEE Internat. Sympos. Inform. Theory*, p. 174, June 2004.
[49] M. Costa, "Writing on dirty paper," *IEEE Trans. Inform. Theory*, pp. 439–41, May 1983.
[50] U. Erez, S. Shamai, and R. Zamir, "Capacity and lattice strategies for cancelling known interference," *Proc. Internat. Sympos. Inform. Theory Appl.*, pp. 681–4, November 2000.
[51] B. Rimoldi and R. Urbanke, "A rate-splitting approach to the Gaussian multiple-access channel," *IEEE Trans. Inform. Theory*, pp. 364–75, March 1996.
[52] D. Tse and P. Viswanath, *Foundations of Wireless Communications*, Cambridge University Press, 2005.
[53] R. Knopp and P. Humblet, "Information capacity and power control in single-cell multiuser communications," *Proc. IEEE Internat. Conf. Commun.*, pp. 331–5, June 1995.
[54] N. Jindal, S. Vishwanath, and A. J. Goldsmith, "On the duality of Gaussian multiple-access and broadcast channels," *IEEE Trans. Inform. Theory*, pp. 768–83, May 2004. 502
[55] S. Verdú, "Multiple-access channels with memory with and without frame synchronism," *IEEE Trans. Inform. Theory*, pp. 605–19, May 1989.
[56] E. Telatar, "Capacity of multi-antenna Gaussian channels," *Euro. Trans. Telecommun.*, pp. 585–96, November 1999.
[57] W. Yu, W. Rhee, S. Boyd, and J. Cioffi, "Iterative water-filling for vector multiple access channels," *IEEE Trans. Inform. Theory*, pp. 145–52, January 2004.
[58] R. Cheng and S. Verdú, "Gaussian multiaccess channels with ISI: Capacity region and multiuser water-filling," *IEEE Trans. Inform. Theory*, pp. 773–85, May 1993.
[59] A. J. Goldsmith and M. Effros, "The capacity region of broadcast channels with intersymbol interference and colored Gaussian noise," *IEEE Trans. Inform. Theory*, pp. 219–40, January 2001.
[60] S. Vishwanath, N. Jindal, and A. J. Goldsmith, "Duality, achievable rates, and sum-rate capac-

ity of Gaussian MIMO broadcast channels," *IEEE Trans. Inform. Theory,* pp. 2658–68, October 2003.

[61] P. Viswanath and D. N. C. Tse, "Sum capacity of the vector Gaussian broadcast channel and uplink–downlink duality," *IEEE Trans. Inform. Theory,* pp. 1912–21, August 2003.

[62] N. Jindal, W. Rhee, S. Vishwanath, S. A. Jafar, and A. J. Goldsmith, "Sum power iterative water-filling for multi-antenna Gaussian broadcast channels," *IEEE Trans. Inform. Theory,* pp. 1570–9, April 2005.

[63] P. Viswanath, D. N. C. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Trans. Inform. Theory,* pp. 1277–94, June 2002.

[64] TIA/EIA IS-856, "CDMA 2000: High rate packet data air interface specification," November 2000.

[65] A. Jalali, R. Padovani, and R. Pankaj, "Data throughput of CDMA-HDR a high efficiency–high data rate personal communication wireless system," *Proc. IEEE Veh. Tech. Conf.,* pp. 1854–8, May 2000.

[66] M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, and P. Whiting, "Providing quality of service over a shared wireless link," *IEEE Commun. Mag.,* pp. 150–4, February 2001.

[67] X. Liu, E. K. P. Chong, and N. B. Shroff, "Opportunistic transmission scheduling with resource-sharing constraints in wireless networks," *IEEE J. Sel. Areas Commun.,* pp. 2053–64, October 2001.

[68] D. N. C. Tse, P. Viswanath, and L. Zheng, "Diversity–multiplexing trade-off in multiple-access channels," *IEEE Trans. Inform. Theory,* pp. 1859–74, September 2004.

[69] A. J. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE J. Sel. Areas Commun.,* pp. 684–701, June 2003.

[70] M. Sharif and B. Hassibi, "Scaling laws of sum rate using time-sharing, DPC, and beamforming for MIMO broadcast channels," *Proc. IEEE Internat. Sympos. Inform. Theory,* p. 175, June 2004.

[71] T. Yoo and A. J. Goldsmith, "Optimality of zero-forcing beamforming with multiuser diversity," *Proc. IEEE Internat. Conf. Commun.,* May 2005.

[72] J. Heath, R. W. M. Airy, and A. Paulraj, "Multiuser diversity for MIMO wireless systems with linear receivers," *Proc. Asilomar Conf. Signals, Syst., Comput.,* pp. 1194–9, November 2001.

[73] N. Jindal and A. Goldsmith, "DPC vs. TDMA for MIMO broadcast channels," *IEEE Trans. Inform. Theory,* pp. 1783–94, May 2005.

[74] N. Al-Dhahir, C. Fragouli, A. Stamoulis, W. Younis, and R. Calderbank, "Space-time processing for broadband wireless access," *IEEE Commun. Mag.,* pp. 136–42, September 2002.

[75] M. Brehler and M. K. Varanasi, "Optimum receivers and low-dimensional spreaded modulation for multiuser space-time communications," *IEEE Trans. Inform. Theory,* pp. 901–18, April 2003.

[76] S. N. Diggavi, N. Al-Dhahir, and A. R. Calderbank, "Multiuser joint equalization and decoding of space-time codes," *Proc. IEEE Internat. Conf. Commun.,* pp. 2643–7, May 2003.

[77] H. Dai and H. V. Poor, "Iterative space-time processing for multiuser detection in multipath CDMA channels," *IEEE Trans. Signal Proc.,* pp. 2116–27, September 2002.

503

[78] S. N. Diggavi, N. Al-Dhahir, and A. R. Calderbank, "On interference cancellation and high-rate space-time codes," *Proc. IEEE Internat. Sympos. Inform. Theory,* p. 238, June 2003.

[79] S. J. Grant and J. K. Cavers, "System-wide capacity increase for narrowband cellular systems through multiuser detection and base station diversity arrays," *IEEE Trans. Wireless Commun.,* pp. 2072–82, November 2004.

[80] Z. Liu and G. B. Giannakis, "Space-time block-coded multiple access through frequency-selective fading channels," *IEEE Trans. Commun.,* pp. 1033–44, June 2001.

[81] K.-K. Wong, R. D. Murch, and K. B. Letaief, "Performance enhancement of multiuser MIMO wireless communication systems," *IEEE Trans. Commun.,* pp. 1960–70, December 2002.

504

第15章　蜂窝系统和架构式无线网络

架构式无线网络在给定区域内部署了基站（接入点）。基站提供移动终端到有线骨干网的接口，并完成网络控制的功能，基站之间往往互连以实现协调控制。第16章所述无线自组织网络与此相反，它没有这样的基础架构。架构式无线网络例如蜂窝电话系统、无线局域网和寻呼系统。基站的组织协调提供了传输调度、动态资源分配、功率控制和切换的中心控制机制，同时也能更好地利用网络资源来满足不同用户的性能要求。此外，大部分架构式无线网络的设计中移动终端直接与基站通信，不需要经由中间无线节点的多跳路由。单跳路由一般有更低的时延和损耗、更高的数据速率、也更为灵活。这些原因使架构式无线网络一般要比无架构网络性能好很多。不过架构式网络有时成本很高，有时部署设备也存在实际困难，在这些情况下自组织网络虽然性能较差，但却是最好的选择。

蜂窝系统是架构式网络的一种，它通过不同空间位置上的频率复用提高了频谱效率。本章主要讨论蜂窝系统的设计和分析，其基本原理适用于任何架构式无线网络。本章首先介绍蜂窝系统和信道复用的基本原理，然后讨论系统容量问题以及通过降低干扰来提高容量的方法。我们还将讨论动态资源分配对性能的改善。本章最后分析香农容量和面积频谱效率意义下的速率极限问题。

505

15.1　蜂窝系统的基本概念

蜂窝系统利用了信号功率随传播距离衰减的特点，在不同的空间位置上重复使用频率。它把一个空间区域（例如一个城市）划分为若干个互不重叠的小区（图15-1），把系统的信号维按14.2节所讲的正交或非正交技术分成信道。本章中，正交信道主要考虑TDMA和FDMA，非正交信道主要考虑CDMA。再将这样得到的信道指配给不同的信道集合C_n，给每个小区分配一个信道集，不同的小区可以重复使用相同的信道集，如图15-1所示。这就是*频率复用*（frequency reuse），也称为*信道复用*（channel reuse）。使用相同信道集的小区称为*同信道小区*（co-channel cell）。这些小区之间要有足够的分隔距离，使它们之间的相互干扰不会使信号质量低于要求，具体要分开多少与所用的信道化技术、传播特性、以及用户要求的性能有关。

在图15-1所示的蜂窝系统中，基站在每个小区的中心附近。移动台和哪个基站通信在实际当中是根据SINR来选择的，不过在理想传播条件下，落在某个小区范围内的移动台将同该小区的基站通信。当移动台在两个小区间移动时，必须能把它的通话从原小区的基站*切换*（handoff）到新小区的基站。小区内基站到移动台的信道是下行信道，移动台到基站的信道是上行信道。一定区域内的所有基站都连接到一个移动电话交换局。移动电话局起中央控制的作用，由它来协调组织用户鉴权、信道分配和基站间的切换。当移动台在小区间移动，或者当移动台受到衰落或阴影的影响时，一旦它发到基站的信号质量低于一定门限值，切换将会被启动。如果邻近的基站都没有质量符合要求的空闲信道，切换就会失败，通话将被中止。

506

PDG

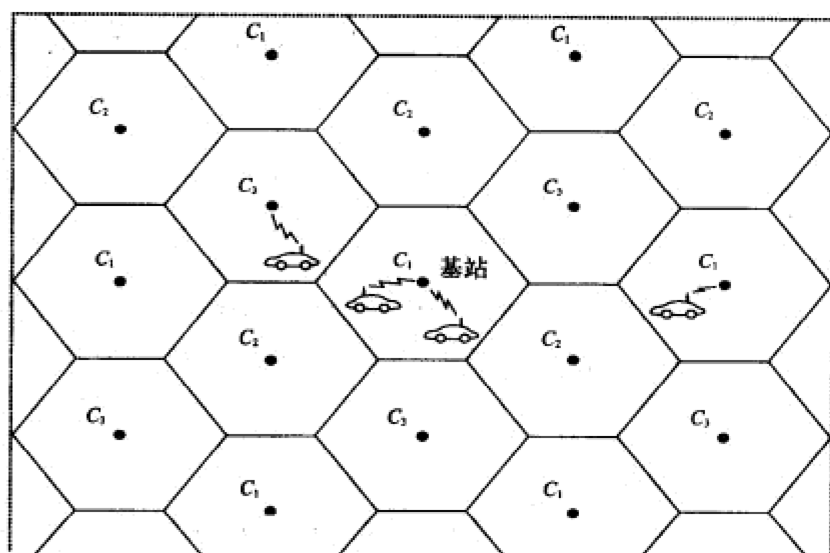


图15-1 蜂窝系统

多址技术对蜂窝系统的上下行来说都是必需的。蜂窝系统中主要使用的多址技术有时分、频分、正交或非正交的码分，以及这些方式的组合，有时会结合空分多址。14.2节已经讨论了单小区时上行和下行的设计问题，其基本设计原则和分析也适用于蜂窝系统，只需要把信道复用的影响考虑进来。多址技术的选择对蜂窝系统和单小区系统是不同的，因为蜂窝系统中的多址技术需要考虑来自小区外的干扰，称为小区间（intercell）干扰或同信道（co-channel）干扰。非正交多址还得考虑来自同一小区的干扰，称为小区内（intracell）干扰。正交多址也会因为多径、同步误差和其他影响正交性的因素而出现小区内干扰。

非正交码的CDMA由于其自身特性同时存在小区内干扰和小区间干扰，但这些干扰都能被扩频码的互相关性抑制。正交多址虽然在理想运行时无小区内干扰，但不能像扩频系统那样能靠处理增益来抑制小区间干扰。用户受小区内和小区间干扰的程度可用信号对干扰加噪声的功率比来表示，定义为：

$$\mathrm{SINR}=\frac{P_r}{N_0B+P_I} \tag{15-1}$$

其中P_r是接收信号功率，P_I是对应小区内和小区间干扰的接收功率。对于CDMA系统，P_I是解扩后的干扰功率。一般用SINR替代信噪比来计算移动台的误码率，这种做法在随机干扰不具有高斯统计特性时有一些误差。

较大的小区间干扰会减小SINR，但会增加用户的误码率。把同信道的小区分隔足够远可以降低小区间干扰，但为了使系统可容纳的用户数达到最大，又必须要尽可能多地增加信道复用此数。因此，最佳的设计就是让同信道的用户所分隔的距离正好能使干扰的大小满足目标误码率所要求的程度。好的蜂窝系统设计是干扰受限（interference limited）的，意思是说系统中的干扰功率大大超过噪声功率。因此，研究这些系统时噪声一般可以忽略，使SINR简化为信号对干扰的功率比SIR = P_r/P_I。在干扰受限的系统中，

SIR决定了用户的误码率，系统可容纳的用户数也因此受限于用户之间干扰。多天线、多用户检测等能降低干扰的技术可以提高SIR，从而能提高给定误码率要求时系统可容纳用户数。需要注意的是，SIR及误比特率对于语音等连续业务其定义十分明确，但对于数据类业务，传输的突发性将使系统设计更为复杂。 507

小区大小也是蜂窝系统设计中需要考虑的一个重要因素。缩减小区半径，若同时能使整个系统的各个方面都按比例缩小以保持每个用户的SINR不变，就可以提高系统可容纳的用户数。考虑图15-2，其中的大小区包含了19个小小区。假设按大小区所设计的蜂窝系统中，每个小区可容纳K个用户。通过降低发送功率可把大小区缩小成图中的小小区。假设包括传播损耗在内的一切也都相应缩减使得SINR不变，那么这个小小区也能容纳K个用户。这样，缩减后的系统能在原有大小区的范围内容纳19K个用户，容量增加了18倍。不过当小区缩小时，传播特性也会跟着变，因此实际情况不会是理想地把整个系统按比例缩小。小区缩小也增加了切换的频度，如果切换失败的比率不变的话，掉话率就会增加。小区缩小同时也会增加骨干网的负载，而且小区越多要求的基站也越多，相应地成本也将增加。因此，虽然小尺寸小区能增加系统容量，他们也有自己的缺点。一种折中的方案是在大小区内的高业务量的区域嵌入热点小小区，以这种方式来增加系统得容量（见文献[1]的3.7节）。

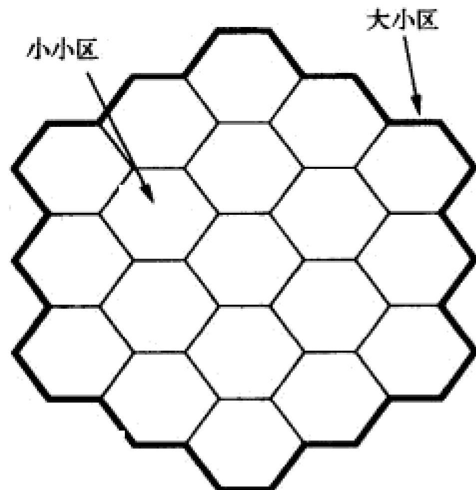


图15-2 容量随着小区的缩小而增加

图15-1中的小区形状是正六边形，这是一种镶嵌形状，它能无重叠铺满整个区域而不留缝隙，矩形、正方形、菱形和三角形也能做到这一点。这些规则的小区形状可以近似基站周围接收信号的等功率线。如果传播特性服从自由空间或者简化路径损耗模型，则等功率线呈圆形，正六边形只是一种合理的近似。第一代蜂窝电话系统常用正六边形来近似小区形状，这些系统的基站设置在高大建筑物的顶部，覆盖范围达几平方英里。现在的系统采用更小的小区，基站位置接近地面，其等功率线更接近菱形，尤其对于城市内的格状街区[2-3]。更小小区和室内小区的等功率线严重依赖于传播环境，很难用镶嵌形状来近似[4]。 508

15.2 信道复用

信道复用是蜂窝系统设计的一个关键因素，它直接决定小区间的干扰，并因此决定了系统的容量和性能。在这一点上，正交多址（如TDMA、FDMA、正交CDMA）和非正交多址（非正交CMDA、正交/非正交混合CDMA）有很大的不同。正交多址在理想条件下没有小区内干扰，但是由于同信道干扰的原因，TDMA和FDMA系统中，使用相同信道的小区中间要隔开几个小区，因为邻近小区复用的话，干扰就会太大。

而非正交多址同时存在小区内和小区间干扰，但所有的干扰都能被扩频码的互相关性减弱，因而能允许信道在所有个小区内复用。采用正交码的CDMA（一般是CDMA的下行）中也在所有小区中复用码字，这是因为各基站之间的码字传输并不同步，来自不同基站的相同码字到达移动台时存在时间差，于是，码字的自相关将能减弱小区间的干扰，这个自相关有可能比较大。在混合方式中，不同小区有不同的非正交码，用这个非正交码对小区内的正交码进行调制，这样就能用非正交码的处理增益来减弱小区间干扰。WCDMA蜂窝系统采用了这种混合方式[5]。实际的CDMA系统一般都是所有小区复用相同的码字，本章也作此假定，这样复用距离就是1，无需再讨论信道复用的优化问题。

下面介绍信道复用中的基本概念：小区簇和信道分配。在干扰受限系统中，每个用户误码率取决于接收到的SIR：用户的接收信号功率与小区内和小区间干扰功率的比值。期望信号的接收功率、小区间和小区内的干扰功率取决于从期望发射机和干扰发射机到期望接收机之间的信道特性。平均SIR一般只按路径损耗计算，路径损耗模型包括了信号和干扰的阴影衰落中值，将阴影和平衰落处理为围绕路径损耗的的统计变化。

路径损耗是传播距离的函数，因此使用相同信道的那些小区之间的**复用距离**（reuse distance）D就是决定小区间平均干扰的一个重要参数。复用距离定义为使用相同信道集的两个小区的中心距离。它与小区形状、小区半径、同信道小区之间的小区个数有关。给定某个性能指标所要求的平均SINR，可以相应的最小复用距离。本节主要考虑按最小复用距离对蜂窝系统的布局进行规划。

图15-3示出了一定信道复用模式下，正六边形和菱形小区的复用距离，图中标出了分配了相同信道集C_n的小区。两种小区形状下的信道复用模式都基于下面将要详细介绍的小区簇的概念。复用距离D是图中所标的点之间的最小距离，这些点是使用信道集C_n的小区的中心。图中还标出了小区半径R，正六边形小区的半径是小区中心到顶点的距离，菱形小区的半径是小区中心到一条边的中点的距离。

509

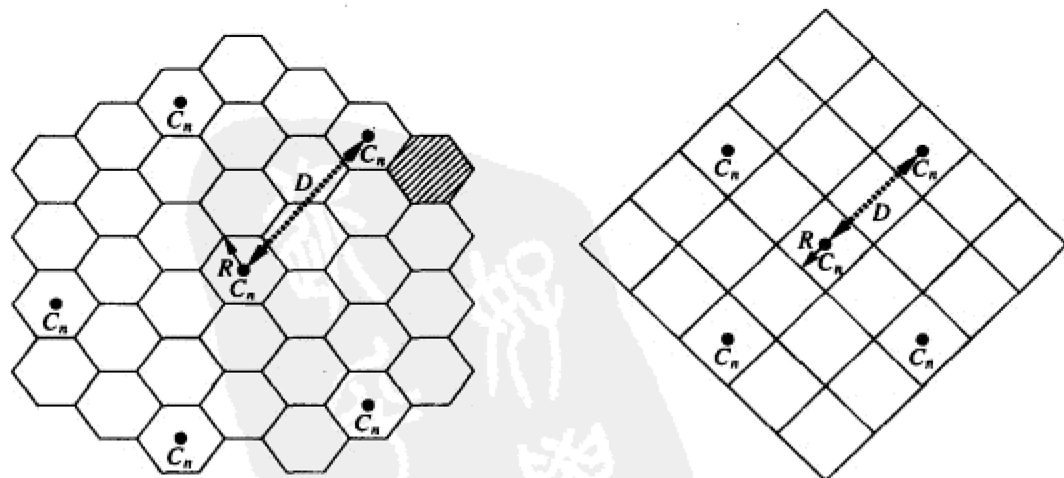


图15-3　给定复用模式下正六边形和菱形小区的复用距离

菱形小区复用距离可直接根据同信道小区之间的小区个数N_i和小区半径R计算。穿越一个小区的距离是$2R$，小区中心到边界的距离是R，穿越N_i个小区的距离是$2RN_i$，所以$D = R + 2RN_i + R = 2R(N_i + 1)$。

正六边形小区复用距离的计算比较复杂，因为两个同信道小区间的小区个数不为整数。例如在图15-3左边，如果信道C_n分给这样两个小区，一个是图中心的那个小区，另一个是图中的阴影小区，那么同信道小区间的小区个数为2，复用距离就容易计算了。但若把C_n分配给阴影小区旁边的那个小区，即如图中实际所标那样时，这两个小区之间的小区个数不再为整数。按图15-3实际所标的方式分配信道需要建立小区簇，后面将详细介绍。

正六边形小区中信道的分配过程是这样的。考虑图15-4，图中R是小区半径。将每个小区的位置用一对数(i, j)来标记，以小区A为原点(0, 0)，从小区A出发，沿u轴移动i个小区，然后逆时针旋转60°，再沿v轴移动j个小区，所到达的小区即为(i, j)。例如小区G的坐标是(0, 1)、S的坐标是(1, 1)、P是(−2, 2)、M是(−1, −1)。容易算出，相邻小区的中距离是$\sqrt{3}R$，坐标为(i, j)的小区和坐标为(0, 0)的小区A的中心距离为

$$D = \sqrt{3}R\sqrt{i^2 + j^2 + ij} \tag{15-2}$$

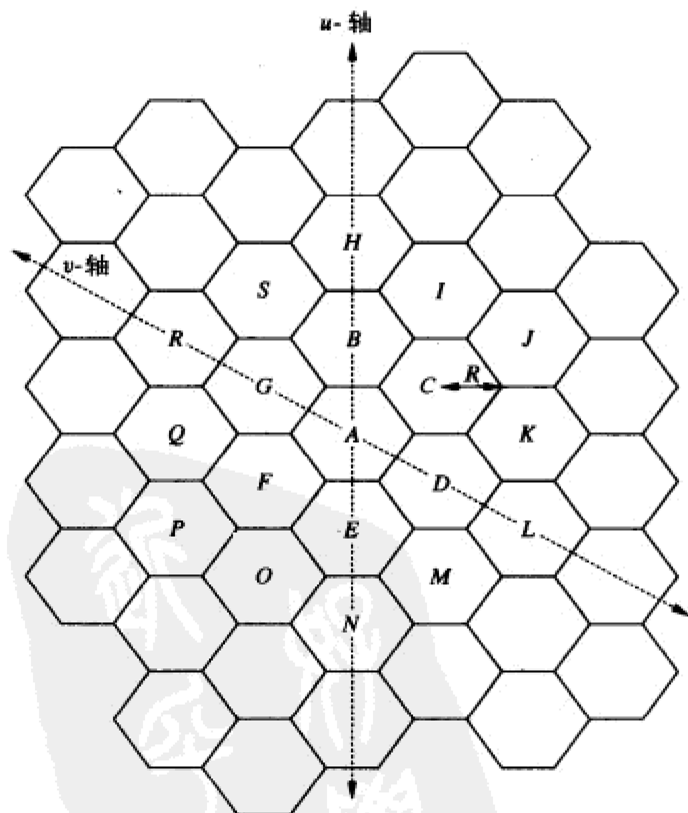


图15-4 正六边形小区复用距离坐标图

例15.1：对于图15-3中的复用模式，求用小区半径R表示的复用距离D。

解：对于菱形小区，同信道小区间有$N_i = 1$个小区，因此$D = 2R(N_i + 1) = 4R$。对于正六边形小区，复用模式是先沿u轴移动2个小区，再沿v轴移动1个小区，因此$D = \sqrt{3}R\sqrt{2^2+1^2+2} = 4.58R$。

510

给定可接受的最小复用距离D_{min}，我们希望在保持最小复用距离的前提下尽可能多地复用信道。这就需要通过重复小区簇（cell cluster）来分配信道。小区簇中的每个小区都分配有不相同的信道集。小区簇必须是一种镶嵌形状以进行空间重复。菱形小区的镶嵌形小区簇也是菱形，每边有K个小区，图15-5中示出了K=4的小区簇。小区簇中分配给第n个小区的信道集为C_n, $n = 1,\cdots,N$，N是不同信道集的个数，信道分配方式对每个小区簇都一样，从而能保证同信道小区的复用距离至少是$D = 2KR$。每个簇中的小区个数是$N = K^2$，也叫做复用因子（reuse factor），由$D = 2KR$可得$N = 0.25(D/R)^2$。若N_c表示每个小区中的信道数，即C_n中的信道数，N_T表示信道总数，那么$N = N_T/N_c$。N越小表明复用效率越高（在给定的区域内信道的复用的次数更多）。但在同时，因为$D = 2KR = 2\sqrt{N}R$，所以N越小复用距离也越小，这将使小区间干扰更大。

511

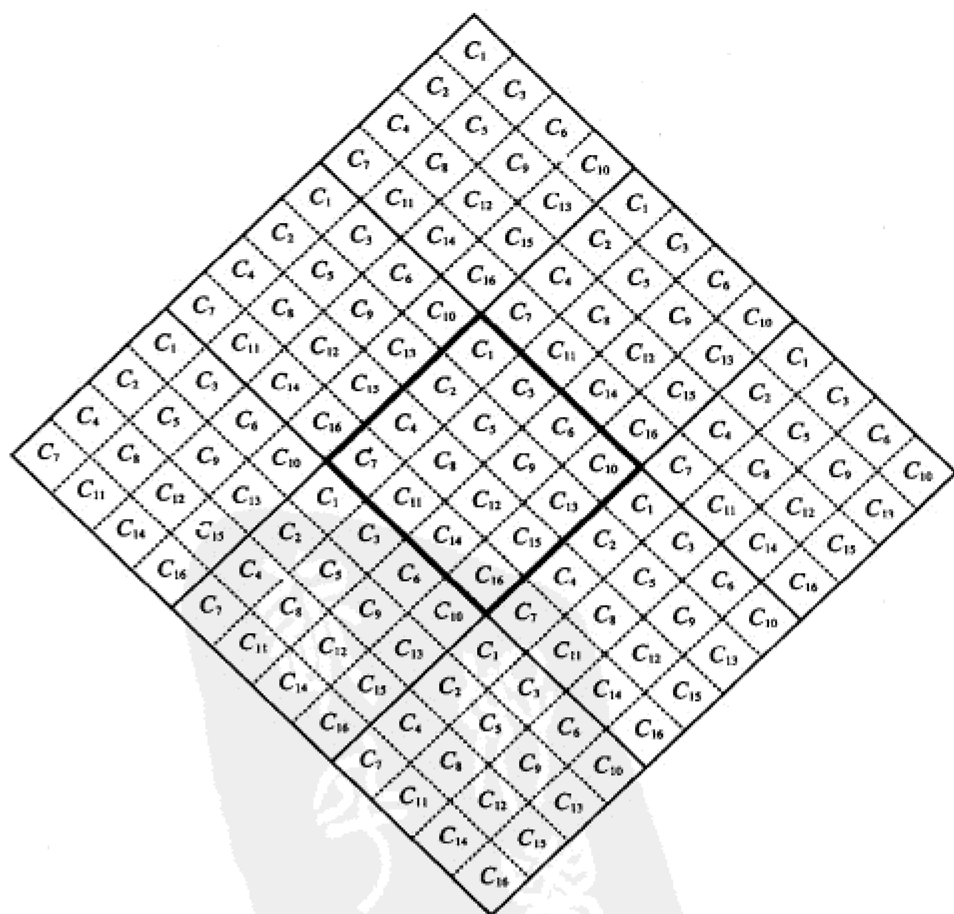


图15-5 菱形小区的小区簇

正六边形小区的小区簇可以迭代形成。先把总带宽分成N个信道集C_1, …, C_N，N是小区簇的大小。分配第一个信道集C_1给任意一个小区，以这个小区为基准，沿着某一排正六边形的方向移动i个小区，逆时针旋转60°，沿这个新方向上的一排正六边形移动j个小区，将信道集C_1分配给所到的这个小区。再回到原始小区，换一个方向重复这个过程，直到换完从原始小区出发的所有方向。图15-6示出了$i = 3$、$j = 2$时的过程。为了在整个区域内分配信道集C_1，再从刚才分配了C_1的任一小区开始，重复上述迭代过程，直至从任何分配了C_1的小区出发都没有未分配的小区时为止。然后，将信道集C_2分配给任意一个未曾分配信道的小区；再按上面的方式在整个区域内分配信道集C_2。重复这个过程直至所有小区都已经被分配了某个信道集，结果就形成了图15-7所示的小区簇。同信道小区间的复用距离是 $D=\sqrt{3}R\sqrt{i^2+j^2+ij}$ 。簇大小可通过簇面积除以小区面积来近似计算。具体就是，正六边形小区的面积是 $A_{\text{cell}}=3\sqrt{3}R^2/2$ ，正六边形小区簇的面积是 $A_{\text{cluster}}=\sqrt{3}D^2/2$ ，因此每个簇包含的小区数是

$$N=\frac{A_{\text{cluster}}}{A_{\text{cell}}}=\frac{\sqrt{3}D^2/2}{3\sqrt{3}R^2/2}=\frac{1}{3}\left(\frac{D}{R}\right)^2=\frac{1}{3}\left(\frac{3R^2(i^2+j^2+ij)}{R^2}\right)=i^2+j^2+ij$$

和菱形小区时一样，N越小表明信道复用效率越高，但同时复用距离 $D=R\sqrt{3N}$ 也越小，小区间干扰也越大。

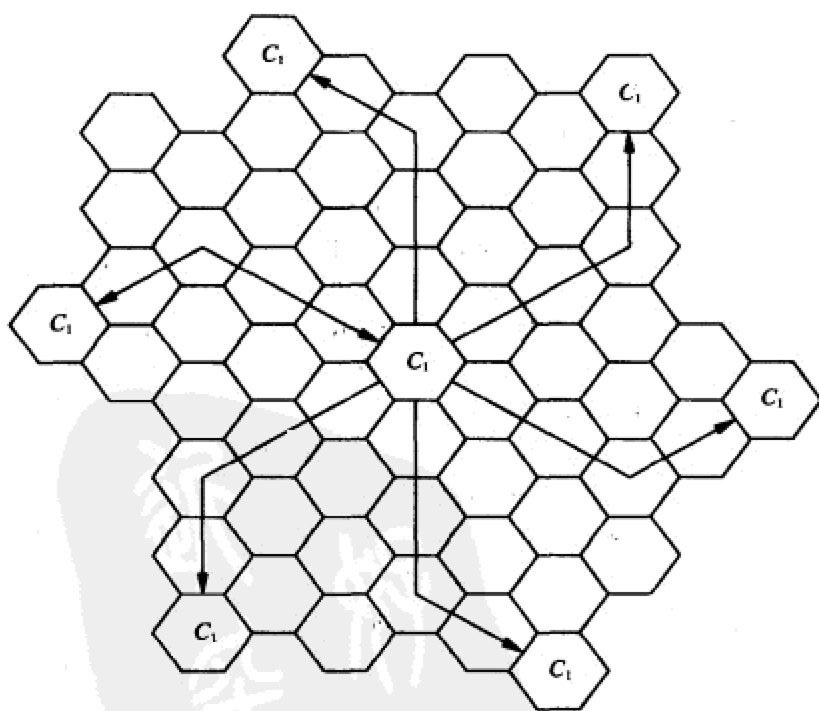
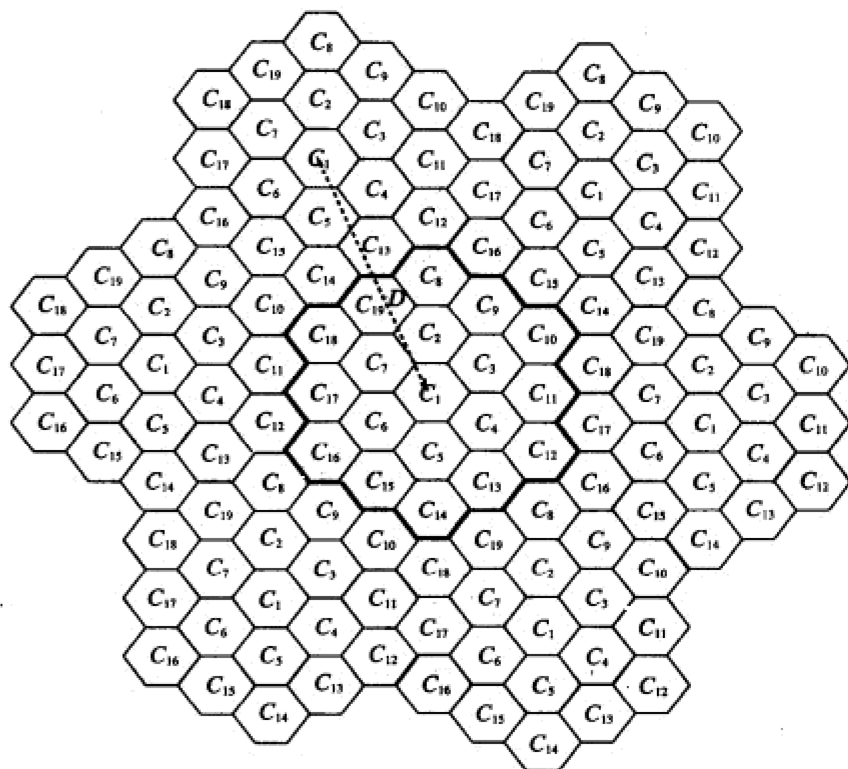


图15-6 正六边形小区的信道分配



512
~
513

图15-7　正六边形小区的小区簇

15.3　SIR和用户容量

本节计算蜂窝系统中用户的SIR和给定目标SIR时每个小区能支持的用户数。我们按照干扰受限的假设忽略噪声对性能的影响，这一计算容易扩展到包括噪声的情况。蜂窝系统的SIR和许多因素有关，包括小区规划、小区大小、复用距离和传播环境等。我们用式（2-39）的简化路径损耗模型来计算路径损耗，取参考距离$d_0 = 1\text{m}$，于是$P_r = P_t L d^{-\gamma}$，其中常量L是$d = d_0$时的平均路径损耗，γ是路径损耗指数。一般情况下L对期望的信号和干扰是相同的，它在SIR中被约掉，因此以下的分析假设$L = 1$。记小区内信号传播的路径损耗指数为γ_I，小区外（小区之间）信号传播的路径损耗指数为γ_O。这两个指数有可能不一样，取决于传播环境和小区的大小[6]。我们用简化路径损耗模型求出正交多址和非正交多址的SIR表达式，再得到相应的用户容量。用户容量定义为满足目标SIR的条件下，系统每小区能支持的最大的用户数。

SIR一般用于计算误码率，一般将干扰近似为加性白高斯噪声，再用误码率与信噪比的关系式来计算误码率。比如，根据式（6-6），未编码BPSK无衰落时误码率是$P_b = Q\left(\sqrt{2\text{SIR}}\right)$，在高SIR的瑞利衰落时的误码率根据式（6-58）是$\overline{P}_b = 0.25/\text{SIR}$。虽然下面的SIR公式中假定了简化的路径损耗模型，但也可以采用更复杂的路径损耗模型以得到更准确的SIR近似值。不过，模型中的许多不精确因素是很难确定的。比如对干扰进行高斯近似在干扰数目大的情况下是准确的，如CDMA系统，但对于干扰数目小的

系统，如TDMA和FDMA，这种近似就不准确了。此外，衰落环境下的性能计算往往忽略了干扰中的衰落。SIR实际是两个随机变量的比值，其分布很复杂，瑞利分布或者其他通用的分布都不能很好地近似它[7]。在准确的路径损耗、阴影效应和多径衰落模型下，获得平均SIR及其分布是非常复杂的，一般通过仿真来获得[6]。

15.3.1 正交系统（TDMA/FDMA）

本节计算正交多址蜂窝系统的SIR和用户容量。这些系统没有小区内干扰，SIR取决于接收信号的功率和来自同信道小区的干扰。在简化路径损耗模型下，若移动台距离基站为d，则上下行接收功率都是$P_r = P_t d^{-\gamma_I}$。平均的小区间干扰与小区外的干扰数有关，简便起见，只考虑第一环上的M个干扰小区，忽略此环以外的干扰。这样的近似在路径损耗指数γ_0较大时比较准确，因为更外面的干扰的路径损耗要比第一环大得多。假设所有的发射功率都是P_t，功率控制的影响问题放到15.3.3节中讨论。若用户与基站的距离为$d < R$，M个干扰源与目标接收机（下行时为移动台，上行时为基站）的距离为d_i，$i = 1, \cdots, M$，那么SIR为

514

$$\mathrm{SIR} = \frac{d^{-\gamma_I}}{\sum_{i=1}^{M} d_i^{-\gamma_0}} \tag{15-3}$$

此SIR是两个随机变量的比值，其分布非常复杂。人们在一些传播模型下研究了分母中干扰的统计特性，当这些干扰为对数正态分布时，其和也是对数正态分布（见文献[8]的第3章）。上下行链路的平均SIR大致相同，不过下行的干扰来自位于每个小区中心的基站，上行最坏的情形下，干扰源可能处在离被干扰基站最近的小区边界上，因此上行SIR的最坏值比下行SIR差。

如果移动台位于小区边界（$d = R$），并且所有的干扰源与目标接收机的距离都是复用距离D，那么SIR表达式可简化为

$$\mathrm{SIR} = \frac{R^{-\gamma_I}}{MD^{-\gamma_0}} \tag{15-4}$$

若$\gamma_I = \gamma_0 = \gamma$，则进一步简化为

$$\mathrm{SIR} = \frac{1}{M}\left(\frac{D}{R}\right)^{\gamma} \tag{15-5}$$

多数小区形状下，D/R是复用因子N的函数，所以可用N来表示信干比。例如对于图15-5中的菱形小区，$M = 8$，$D/R = \sqrt{4N}$，代入式（15-5）有$\mathrm{SIR} = 0.125(4N)^{\gamma/2}$。对于图15-7中的正六边形小区，$M = 6$，$D/R = \sqrt{3N}$，$\mathrm{SIR} = 0.167(3N)^{\gamma/2}$。这两个SIR表达式可统一表示为

$$\mathrm{SIR} = a_1(a_2 N)^{\gamma/2} \tag{15-6}$$

对菱形小区a_1=0.125、a_2=4，对正六边形小区a_1=0.167、a_2=3。可用上式来简单计算达到某个性能要求近似需要的复用因子。设满足目标误码率所需的目标SIR为SIR_0，对式（15-6）求逆可得能满足此目标SIR的最小复用因子为

$$N \geqslant \frac{1}{a_2}\left(\frac{\mathrm{SIR}_0}{a_1}\right)^{2/\gamma} \tag{15-7}$$

根据N和R就可以算出相应的最小复用距离D。当路径损耗指数$\gamma=2$时，式（15-7）简化为$N \geqslant \mathrm{SIR}_0/(a_1a_2)$。考虑阴影衰落时的分析较为复杂，但仍然可以根据某一中断率指标下的SIR计算出最小复用因子[9]。

515

用户容量C_u定义为系统能够同时支持的激活用户总数，这些用户都能满足一个共同的目标SIR值。为此，正交多址系统的N必须满足式（15-7），若按复用因子N分配信道的结果是每个小区有N_c个信道，则$C_u=N_c$。系统总带宽B可以产生出$N_T=B/B_s$个带宽为B_s的正交信道，复用因子N满足$N=N_T/N_c$，于是有

$$C_u=\frac{N_T}{N}=\frac{B}{NB_s}=\frac{G}{N} \tag{15-8}$$

其中$G=B/B_s$是系统带宽除以单个用户的带宽，N是满足式（15-7）的最小整数值。

例15.2：某TDMA蜂窝系统的小区为正六边形，所有信号传播的路径损耗因子都是$\gamma=2$。求目标SIR为10dB所需要的最小复用因子N。若系统总带宽为20MHz，每用户信号带宽为100kHz，计算相应的用户容量。

解：将$a_1=0.167$和$a_2=3$代入式（15-7）得复用因子为

$$N \geqslant \frac{\mathrm{SIR}_0}{a_1a_2}=\frac{10}{0.5}=20$$

由$G=B/B_s=20\times10^6/(100\times10^3)=200$可得每个小区可容纳的用户数为$C_u=G/N=10$。一般情况下$\gamma>2$，下一个例题将讨论这种情况。

例15.3：某TDMA蜂窝系统的小区为菱形，所有信号传播的路径损耗因子都是$\gamma=4$，采用BPSK调制。假设接收信号为瑞利衰落，用户要求的误码率为$\overline{P}_b=10^{-3}$。若系统是干扰受限的，求能满足此性能要求的最小复用因子N。若系统总带宽为20MHz，每用户信号带宽为100kHz，计算相应的用户容量。

解：把干扰当高斯噪声处理，在瑞利衰落下$\overline{P}_b\approx0.25/\mathrm{SIR}_0$，$\mathrm{SIR}_0$是平均SIR。因此满足给定$\overline{P}_b$的SIR为$\mathrm{SIR}_0=0.25/10^{-3}=250$（约24dB）。将$\mathrm{SIR}_0=250$、$a_1=0.125$、$a_2=4$和$\gamma=4$代入式（15-7）得

$$N \geqslant \frac{1}{4}\sqrt{\frac{250}{0.25}}=11.18$$

所以满足性能要求的复用因子是$N=12$。由于$G=B/B_s=200$，所以每小区可容纳的用户数为$C_u=G/N=16$。注意，把干扰当作为高斯噪声处理只是一个近似，根据中心极限定理，干扰数越多，这个近似越准确。

516

15.3.2　非正交系统（CDMA）

非正交系统中的信道就是码字，这些码字一般都在每个小区复用，复用因子是$N=1$。

非正交系统同时存在小区内干扰和小区间干扰，用户容量是满足SIR目标值时每小区能容纳的最大用户数。我们忽略第一层干扰小区以外的小区间干扰，即忽略与目标小区不相邻的小区引起的干扰。同时假定所有信号的传播都服从简化路径损耗模型，路径损耗指数也相同，这一假设对邻小区干扰一般成立，但最终是否成立还是要看传播环境。

每小区的信道数为$N_c = N_T = C_u$。在CDMA系统中，由于远近效应和码字异步的问题，用户容量受限于上行链路。考虑上行链路，在简化路径损耗模型下，接收信号的功率是$P_r = P_t d^{-\gamma}$，其中d是移动台和基站之间的距离。干扰方面，有N_c-1个异步的小区内干扰，小区间干扰来自M个相邻小区的移动台，共是MN_c个异步干扰。设本小区的第$i(i = 1, \cdots, N_c-1)$个干扰源到上行接收机的距离是d_i，其发射功率为P_i。设外小区的第$j(j = 1, \cdots, MN_c)$个干扰源到上行接收机的距离是d_j，其发射功率为P_j。根据13.4.3节，扩频码的互相关使干扰乘上了一个系数$\xi/(3G)$而减少，G是系统处理增益、ξ是扩频码的一个参数，$1 \leqslant \xi \leqslant 3$。因而，总的小区内和小区间干扰功率为

$$I = \frac{\xi}{3G}\left(\sum_{i=1}^{N_c-1} P_i d_i^{-\gamma} + \sum_{j=1}^{MN_c} P_j d_j^{-\gamma}\right) \tag{15-9}$$

于是SIR为

$$\text{SIR} = \frac{P_t d^{-\gamma}}{\frac{\xi}{3G}\left(\sum_{i=1}^{N_c-1} P_i d_i^{-\gamma} + \sum_{j=1}^{MN_c} P_j d_j^{-\gamma}\right)} \tag{15-10}$$

式（15-10）中的距离各不相同，若无另外的假设此式不能再简化。为此，我们假设小区内有理想功控，使得期望信号和小区内干扰信号的接收功率相同，即$P_r = P_t d^{-\gamma} = P_i d_i^{-\gamma}, \forall i$。再令

$$\lambda = \frac{\sum_{j=1}^{MN_c} P_j d_j^{-\gamma}}{(N_c - 1)P_r} \tag{15-11}$$

表示功控后，外小区干扰和本小区干扰的功率比。于是

$$\text{SIR} = \frac{1}{\frac{\xi}{3G}(N_c - 1)(1 + \lambda)} \tag{15-12}$$

此式常用于带功控的CDMA上行链路中[10-11]。令式（15-12）等于SIR的目标值SIR_0便可得到用户容量$C_u = N_c$为

$$C_u = 1 + \frac{1}{\frac{\xi}{3G}(1 + \lambda)\text{SIR}_0} \tag{15-13}$$

517

语音信号依其统计特性是不需要连续激活的[12]。一个语音用户实际占据信道的时间比例叫做*语音激活因子*（voice activity factor），记为α，$0<\alpha\leqslant 1$。如果发射机在非激活期关闭，那么CDMA中的干扰，即式（15-13）中的分母项需要乘以α，SIR和用户容量都会相应提高。

例15.4：*考虑一个小区内理想功控的CDMA系统。假定目标*SIR_0=10dB，*处理增益*G =

200，扩频码参数$\xi = 2$，并假设来自小区内外的平均干扰功率相等（$\lambda = 1$）。计算系统的用户容量。

解：由式（15-13）有

$$C_u = 1 + \frac{1}{\frac{2}{600}(2 \times 10)} = 16$$

即每个小区能容纳16个用户。

既然式（15-13）和式（15-8）给出了用户容量的简单公式，我们自然想通过它们来比较给定SIR目标值下，TDMA和CDMA谁能支持更多的用户。例15.2和例15.4已经做了这种比较，可以看出在目标SIR和其他系统参数相同的条件下，TDMA系统每小区的用户数为10，而CDMA系统为16。不过所用的模型及系统假设对这些容量公式影响很大。在例15.4中λ从1增加到2，CDMA系统的容量C_u将从16减小到11；例15.2中路径损耗因子γ从2增加到3，复用因子N将从20减小到6，从而TDMA系统的容量将从10增加到33。CDMA系统能在扩频和编码之间进行折中，牺牲一些处理增益能换来高编码增益，从而能降低SIR目标值[13]。但TDMA系统不能把扩频增益转化为编码增益，因此难以获得高编码增益。CDMA系统中没有考虑语音激活，考虑这一因素后容量可以提高。另外，CDMA的容量计算中假设了理想的信道反转功控，而TDMA中没有考虑功控。阴影和衰落的影响也没有考虑。由于CDMA中的信道反转功控，衰落会造成一些功率损失，衰落也会对TDMA和CDMA的小区间干扰功率产生影响。所有这些因素使得我们难以对哪种技术容量更大做出一般性结论。文献[11, 14~16]按照运营系统中的各种假设和模型对TDMA和CDMA的用户容量进行了分析。

15.4 抗干扰技术

蜂窝系统是干扰受限系统，任何能减小干扰的技术都能增加SIR和用户容量。本节讨论蜂窝系统中的抗干扰技术，包括扇区化、智能天线、干扰平均、多用户检测和预干扰消除。

518

减小干扰的常用技术是扇区化。扇区化利用定向天线将基站的360°全向天线分成为N个扇区，图15-8示出了$N = 8$的情形。由图可见，小区内和小区间干扰主要来自同一扇形。当负载很大（每个扇形中都存在干扰源）时，扇区化可以将干扰功率减小为$1/N$。当移动台在扇区间移动时需进行切换，它会使系统开销及掉话率增大。蜂窝系统一般都采用$N = 3$的扇区化。

智能天线一般由结合空时信号处理的天线阵列组成。智能天线用高增益的窄波束对准期望用户，它也能在干扰方向上产生零陷[17]。天线阵列也可以用来提供复用增益，使系统达到更高的数据速率，或者用来提供分集增益，使系统能提高可靠性。这些问题14.9节已做讨论。蜂窝系统设计时需要权衡考虑是用多天线来减小干扰，还是用它来提高其他性能。

519

本小区上行干扰中，小区间干扰主要来自一两个靠近本小区边缘的邻小区的移动台。对于TDMA或者FDMA系统，在原有的时分或频分信道化之上叠加FH（跳频）能减小

这种最坏情形的干扰。此时，所有移动台都按其唯一的跳频图案改变载波频率。仅当干扰和期望用户的跳频模式在时间和频率上重叠时，才会对期望信号引起小区间干扰。上述最坏情况下的干扰源和期望用户有不同的跳频图案，它们跳到一起的机会很少。这样，跳频就起到了干扰平均的作用，它使来自某个小区的干扰平均到小区内的所有位置，从而能大大减轻小区边界处移动台引起的干扰问题。GSM系统也因此而采用了跳频。

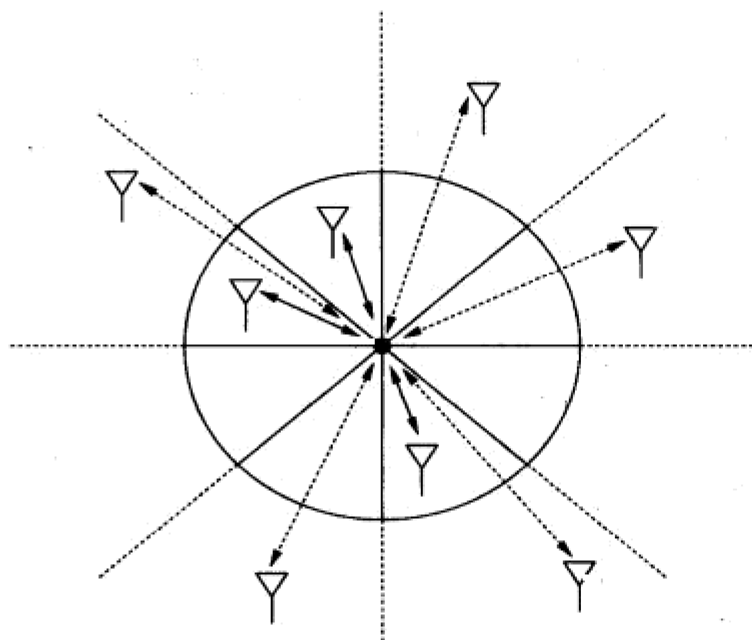


图15-8 $N=8$时圆形小区的扇区化

多用户检测也可以降低干扰[18]。它把期望信号和某些或全部干扰信号一起检测，检测出的干扰可以被减弱或消除。使用多用户检测需要权衡性能增益和检测复杂性，特别在干扰源很多的时候。13.4节已经讨论了单小区CDMA系统上行链路的多用户检测方法。这些方法也可用于小区间的干扰信号。上行链路的基站复杂度限制不大，可同时检测本小区及外小区的干扰[19]，而下行链路的移动台一般只检测少数主要的干扰[20]。

干扰预消除利用了这样一个事实，即下行链路中基站已知本小区内两用户间的干扰，同时已知它对其他小区移动台的干扰。基站利用这一点预先消除用户间的干扰[21~23]。其根本是采用Costa提出的能达到容量极限的“脏纸编码”[24]技术来进行干扰预消除（见14.4.5节）。文献[21~23]中的数值结果表明，当基站有大量天线时，干扰预消除技术能使蜂窝系统的容量增加一个数量级。但是，预消除技术需要发射机预知CSI，而多用户检测并不需要。

15.5 动态资源分配

蜂窝系统中任意小区的用户数量以及用户的信道增益都是动态变化的。蜂窝系统的业务也已从主流的语音业务开始向多媒体数据业务倾斜，用户不再只有单一的数据速率

要求。为了支持不同的业务类型，对用户的资源分配必须更为灵活。动态资源分配根据当前系统的状态和用户需求，动态地决定信道分配、数据速率和发送功率。蜂窝系统动态资源分配方面已经有了大量的研究工作。本节总结一些主要的技术，包括调度、动态信道分配和功率控制，所列的参考文献只是这方面的一小部分。

15.5.1 调度

调度的基本思想是根据移动用户的速率要求和延迟限制动态地分配资源。这些要求也称为服务质量（Quality-of-Service，QoS）。调度对资源的分配必须是高效和公平的。如14.8节所论，通常是这两个目标之间的折中。效率最高的资源分配方法是采用多用户分集把资源分配给信道最好的用户。但这需要假定信道最好的用户能充分利用这些资源，而有些情况下，用户可能不需要很高的速数据速率。此外，这样的分配对信道条件差的用户不公平。

520

人们已经研究了CDMA和TDMA系统、上行和下行链路中的调度问题。注意许多CDMA蜂窝系统在下行传输中采用了一种称为*高速率数据*（high data rate，HDR）的TDMA方式（见文献[1]的2.2节），这些CDMA的调度机制是基于TDMA信道化的。文献[25]针对兼容HDR的TDMA下行链路，比较了三种调度方法：轮循、等时延和相对公平。文献[26]研究了采用多用户分集的TDMA下行链路调度问题。文献[27~29]研究了CDMA下行链路中的一些调度问题，包括速率自适应、公平性和截止时限。文献[30~31]研究了CDMA上行链路的调度算法，其研究假设采用单用户匹配滤波器检测，文献[32]则分析了多用户检测带来的改善。MIMO使系统资源的调度增加了一个自由度，详见文献[33~34]。通过适当的调度算法，也可支持多种类型的用户，详见文献[35]。

15.5.2 动态信道分配

动态信道分配（Dynamic channel assignment，DCA）分为两类：小区内多信道动态分配和小区间正交信道的动态分配。CDMA系统的信道在所有小区中复用，不存在小区间DCA的问题。小区间DCA没有固定的复用模式，只要每个用户的SIR都能满足要求，每个信道都能在每个小区内使用。信道按需分配给用户，通话结束后再还给可用信道集以备用。与固定的复用模式相比，即使采用相对简单的算法，小区间DCA也能使复用效率提高一倍或更多[36, 37]。在数学上，小区间DCA是一个以信道复用满足SIR要求为约束的组合优化问题。多数动态信道分配算法都假设除呼叫到达和离开时间外，其余系统参数都是固定的[38~40]，信道复用限制在此时间内是一个固定的连通图。这样表述的信道分配问题其实是顶点着色问题的一般化，因此是NP难题[41]。低复杂度的解法包括神经网络法[42]和模拟退火法[43]，但有不收敛的问题。DCA在轻负载时效果突出[38]，重负载时，DCA中出现的次优分配难以被重新分配。用户的移动性也会影响小区间DCA的性能，它将引起频繁的信道重分配，相应也会使掉话率增加[44]。最后，小区间DCA的复杂性限制了这项技术的实用，特别对于小蜂窝系统和快速变化的传播环境条件。

521

小区内DCA在一个小区内给某个用户动态分配多个信道。对TDMA系统就是给一

个用户分配多个时隙，对CDMA系统就是给一个用户分配多个码字或不同的扩频因子。对于TDMA或正交CDMA系统，分配多个信道给一个用户比较简单，并不改变所分配信道的特性。文献[45~46]用动态时隙分配来减少数据业务的时延，同时保证满足语音用户的要求。与此不同的是，在非正交CDMA中，不论是通过多码还是通过可变扩频因子，小区内DCA都会引起信道性能的恶化，比如采用多码时，如果每个信道都使用单用户匹配滤波检测器，这些码字之间将会出现自干扰。虽然这不见得比把不同的码字分配给不同的用户更差，但对同一用户分配多个码字明显是次优的。对于可变扩频因子的CDMA，低扩频因子能提供高数据速率，因为它使总的系统带宽中，用在扩频功能上的带宽减少，用于数据传输的带宽增加。但减小扩频因子也使信号对其他用户的干扰更敏感。文献[47~48]分析了采用可变扩频增益的小区内DCA，文献[49]研究了多码技术。文献[50]给出了CDMA系统中多码DCA和可变扩频DCA的比较，并指出这两种技术在使用单用户匹配滤波检测器时是等效的，但在采用更加复杂的检测技术时，多码要好一些。

15.5.3 功率控制

功率控制是无线系统设计中的一个关键点。从前面的章节可以看到，在衰落信道中，注水法功控可以使自适应速率的单用户和多用户系统的容量最大化。在单用户衰落信道中，信道反转功控能保持固定的接收信噪比，同时也消除了CDMA上行的远近效应。不过在蜂窝系统中，这些功控方式对小区间干扰有不同的影响。

下行功控对小区间干扰的影响比上行功控小，因为所有的下行传输都源自小区中心，而上行传输可能来自小区边缘，功控将恶化对邻近小区的干扰。因此我们主要考虑上行功控的影响。考虑图15-9所示的两个小区。假设小区B内的两个移动台B_1和B_2以相同的功率发射，那么移动台B_1对小区A基站的干扰相对较大，因为B_1接近小区A的边缘，而B_2距离远，所以干扰要弱一些。如果采用注水法功控，那么由于B_1离小区B的基站比B_2远，它的信道增益一般较差，所以B_1的发射功率要比B_2低。这对减轻对小区A的干扰有积极的作用。就是说，注水法功控能降低来自小区边缘的移动台的干扰，而这些移动台正是小区间干扰的主要来源。多用户分集时情况类似，用户仅在到自己基站的信道增益高的时候才发射，这一般是他们接近小区中心的时候。与此相反，信道反转功控将使小区边缘处的移动台有更大的发射功率，以使它到达基站的功率能和小区中心附近的移动台到达基站的功率相同。这将增大小区边缘移动台对邻小区的干扰。 522

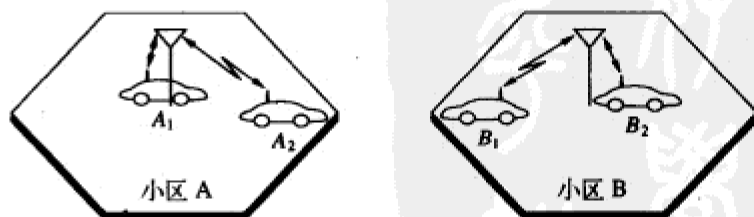


图15-9 功率控制对小区间干扰的影响

14.4节介绍的功控算法能使单小区内的所有用户都满足给定的目标SINR值，这个算法可以扩展到多小区。设有K个用户，用户k要求的SINR是γ_k^*。考虑上行链路，用户k的SINR为

$$\gamma_k = \frac{g_k P_k}{n_k + \rho \sum_{k \neq j} g_{kj} P_j} \quad k, j \in \{1, \cdots, K\} \tag{15-14}$$

其中g_k是用户k到他的基站的信道功率增益，$g_{kj} > 0$是第j个干扰用户（本小区或外小区）到用户k所在基站之间的信道功率增益，P_k是用户k的发射功率，P_j是干扰用户j的发射功率，n_k是用户k所在基站的噪声功率，ρ是信号处理带来的干扰降低量。在CDMA中$\rho \approx 1/G$，在TDMA中$\rho=1$。类似于上行功控的情形，SIR约束可表示为矩阵形式$(\boldsymbol{I}-\boldsymbol{F})\boldsymbol{P} \geqslant \boldsymbol{u}$，$\boldsymbol{P}>\boldsymbol{0}$，其中$\boldsymbol{P}=(P_1, P_2, \cdots, P_K)^T$是各用户的发射功率组成的列向量，$\boldsymbol{u}$是噪声功率乘以目标SIR，除以信道增益后的向量：

$$\boldsymbol{u} = \left(\frac{\gamma_1^* n_1}{g_1}, \frac{\gamma_2^* n_2}{g_2}, \cdots, \frac{\gamma_K^* n_K}{g_K} \right)^T \tag{15-15}$$

$\boldsymbol{F}$不可约，它的每个元素非负，为

$$F_{kj} = \begin{cases} 0 & k = j \\ \gamma_k^* g_{kj} \rho / g_k & k \neq j \end{cases} \tag{15-16}$$

其中$k, j \in (1, 2, \cdots, K)$。如同14.4节，文献[51~53]证明，若矩阵$\boldsymbol{F}$的Perron-Frobenius特征值小于1，那么存在向量$\boldsymbol{P}>0$（即对所有k有$P_k>0$）使所有用户的SIR都能满足要求，且Pareto最优解为$\boldsymbol{P}^* = (\boldsymbol{I}-\boldsymbol{F})^{-1}\boldsymbol{u}$。因此$\boldsymbol{P}^*$能以最小的发射功率满足各用户的SIR要求。此外，分布式迭代功控算法为

523

$$P_k(i+1) = \frac{\gamma_k^*}{\gamma_k(i)} P_k(i) \tag{15-17}$$

它能收敛到最优解。这是一个简单的功控算法，它只需要每个发射机知道SIR信息，当SIR低于目标值时增加发射功率，当SIR超过目标值时减小发射功率。不过需要特别注意的是，与单小区上行链路相比，蜂窝系统中满足所有SIR要求的功率向量的解存在的可能性较小，这是因为影响SIR的干扰源更多，并且信道增益的取值范围也更大。当这样的解不存在时，分布式算法将使所有用户以最大功率发射，并且仍不能满足SIR要求。

功率控制常常和调度、小区间DCA和小区内DCA结合运用。文献[54~55]分析了TDMA中的小区间DCA和功率控制，文献[56]将其扩展到MIMO系统。结合了多用户分集、功率控制和小区内DCA的研究见文献[57~58]。CDMA中结合功控和可变扩频小区内DCA的研究见文献[59~60]。功控结合多码或结合可变扩频的比较见文献[61]。

15.6 基本速率极限

15.6.1 蜂窝系统的香农容量

信息论方面对于蜂窝系统的香农容量的研究结果还很少，主要困难在于要将信道复

用以及它产生的干扰结合到基本容量的分析之中。对于独立的单个小区，14.5节和14.6节已经给出了上行和下行的容量分析，进一步往多小区扩展的工作还少见。某些情况下已经能对容量做出描述，但对于信道模型、基站协作、干扰特性和发射机或者接收机CSI等方面的广泛假设下，容量以及相应的最优发送和接收方案还未见解决。

一种分析蜂窝系统容量的方法是假定所有基站都可以完全协作，能对所有信号进行联合编码和译码。此时的容量分析并没有小区的概念，多个基站可以看成是一个由分布在不同地理位置上的天线组成的多天线的基站。从多天线基站到移动台的传输是MIMO下行信道（广播信道），移动台到多天线基站的传输是MIMO上行信道（多址信道）。在一定的信道模型和CSI假设下，这两种信道的香农容量域是已知的，如14.9节所述。

文献[62]最早在基站可完全协作的假设下研究了蜂窝系统上行链路的容量，文献[63]随后进行了更全面的分析。这些研究中将移动台到基站的传播环境建模为AWGN信道，小区内的信道增益为1，小区间的增益为α，$0\leqslant\alpha\leqslant1$。文献[63]中的Wyner模型按一维和二维小区阵列推导出了每用户的容量，其定义为所有用户都能同时维持的最大可能速率，即

$$C(\alpha)=\frac{B}{K}\int_0^1\log_2\left(1+\frac{KP(1+2\alpha\cos(2\pi\theta))^2}{N_0B}\right)\mathrm{d}\theta \tag{15-18}$$

524

其中B是系统带宽，$N_0/2$是噪声功率谱密度，K是每小区的用户数，P是每个用户的平均发射功率。文献[63]和[62]都指出，达到上行容量的方法是在每个小区内使用正交多址（如TDMA），并在其他小区复用这些正交信道，虽然这不一定是唯一的最佳方法。容量与小区间干扰衰减因子α的关系$C(\alpha)$与系统的信噪比有关。高信噪比时，$C(\alpha)$随着α的增加而增加，这是因为强干扰有助于检测出并抵消掉干扰。低信噪比时，$C(\alpha)$先随α的增加而减小，之后随α的增加而增加。这是因为小区间的弱干扰不能被检测出并抵消掉，容量就相应减小了。而当小区间干扰的信道增益增加时，联合检测能更好地译出并减掉干扰，容量相应提高。

分析蜂窝系统容量的另一种方法是假定基站之间完全不协作，将来自其他小区的信号都作为干扰。这种方法体现实际的蜂窝系统设计。不过，干扰信道的香农容量除了一些强干扰特例外[66]，还是信息论中长期未解决的问题[64-65]。若将干扰近似为高斯噪声，则上下行的容量都可以用14.5节和14.6节中的单小区容量分析来解决。高斯假设是一种最坏的假设，因为利用已知的干扰结构有可能帮助期望信号的检测，从而能提高容量。对一维和二维小区网格，文献[67]将干扰作为高斯噪声，得到了有衰落的蜂窝系统的上行容量。这些结果都表明，无论是否有衰落，当小区间干扰不可忽略时，在小区内使用正交多址（如TDMA）是最佳的。这个一般性的结论对小区内采用信道反转功控的情形也适用。此外，在某些情况下，限制或不使用信道复用能够增加系统容量。当基站间能部分协作时，这种模型对容量的影响见文献[68]。

上述结果对蜂窝系统上行链路的容量及其最佳发送策略提供了一些实质性认识。然而对于下行链路，关于信道模型或者基站协作的任何假设下，却还没有这样的结论。虽然对于提供双向话音的蜂窝系统来说，上行链路是容量的瓶颈，但对于多媒体下载，下

行链路正变得越来越重要。因此，对蜂窝系统下行的容量限制及对其实质的更好理解对于未来蜂窝系统的设计是非常有益的。

15.6.2　面积频谱效率

上一节所讨论的蜂窝系统的香农容量域是上行或下行链路中最大可达速率的集合。计算容量域时，若假设基站可以协作，则容量域只是一个具有一个多天线基站的小区，而当我们把小区间干扰看成是高斯噪声的时候，上下行的容量域与蜂窝系统的结构密切相关，如小区半径和信道复用距离。**面积频谱效率**（Area spectral efficiency，ASE）是一个容量度量，它能允许我们根据基本容量极限优化蜂窝系统的结构，例如复用距离。

525

无论信道化是正交的还是非正交的，蜂窝系统的复用距离D都指任意两个使用相同信道的小区的中心距离。这些信道在距离D处复用，每个信道所覆盖的面积大约是半径为$0.5D$的圆面积$A=\pi(0.5D)^2$。复用距离越大，信道复用的效率越低。但减小复用距离会增加小区间干扰，当我们把干扰看作噪声时，就会减小每个小区的容量域。ASE则能折中资源利用效率和每个蜂窝的容量域。

假定一个蜂窝系统中，每个小区有K个用户，复用距离为D，总带宽为B。令$C=(R_1, R_2, \cdots, R_K)$表示把小区间干扰当作高斯噪声处理时，给定小区内上行或下行的容量域。相应的总速率，也即系统吞吐量为

$$C_{\mathrm{SR}}=\max_{(R_1,\ldots,R_K)\in C}\sum_{k=1}^{K}R_k\ \text{bit/s} \tag{15-19}$$

容量域C和相应的和速率C_{SR}可以通过任意的小区内信道化技术来得到。显然，这个容量域将随着小区间干扰的增加而减小。此外，由于小区间干扰随着复用距离的增加而减小，所以容量域的大小随复用距离而增加。

小区的ASE定义为小区的资源每单位面积每赫兹带宽所能支持的吞吐量。给定上述的和速率容量，ASE就是

$$A_e \triangleq \frac{C_{\mathrm{SR}}/B}{\pi(0.5D)^2}\ \text{bit/s/Hz/m}^2 \tag{15-20}$$

根据文献[67]，正交信道化能达到小区容量，所以我们用TDMA来计算ASE。再假设系统是干扰受限的，则噪声可忽略，于是小区内每个用户的速率R_k是其接收信干比$\gamma_k=P_k/I_k, k=1,\cdots,K$的函数。如果$\gamma_k$固定，则$R_k=\tau_k B\log(1+\gamma_k)$，其中$\tau_k$是分配给用户$k$的时段。$\gamma_k$一般不固定，因为用户和干扰的功率都随传播环境及移动台位置而改变。对于时变的γ_k，可按14.5.4节和14.6.2节那样，通过沿时间和沿用户的最优资源分配来得到容量域。

526

作为一个简单的例子，考虑一个小区是半径为R的正六边形的AWGN TDMA上行链路。假设小区内所有用户都分配了相同的时间比例$\tau_k=1/K$，发射功率也同为P。只考虑第一圈的干扰小区，忽略此外的干扰小区即只考虑6个小区间干扰。考虑这样一个非常保守的模型：(i) 指定小区内所有用户都位于小区的边界；(ii) 干扰小区中的用户也都位于所在小区的边界，并且在离指定小区最近的位置上。路径损耗采用简化模型，小

PDG

区内是$P_r = P_t d^{-2}$，小区间是$P_r = P_t d^{-\gamma}$，其中$2 \leqslant \gamma \leqslant 4$。那么用户$k$的接收功率是$P_r = PR^{-2}$，小区间的干扰功率是$I_k = 6P(D-R)^{-\gamma}$。于是，用户$k$可达到的最大速率为

$$R_k = \frac{B}{K}\log_2\left(1 + \frac{(D-R)^\gamma}{6R^2}\right) \text{ bit/s} \tag{15-21}$$

ASE为

$$A_e = \frac{\sum_{k=1}^{K} R_k/B}{\pi(0.5D)^2} = \frac{\log_2(1+(D-R)^\gamma/6R^2)}{\pi(0.5D)^2} \text{ bit/s/Hz/m}^2 \tag{15-22}$$

图15-10画出了$\gamma = 2$和$\gamma = 4$时，A_e和D的关系，其中小区半径归一化为$R = 1$。可见，如果小区间干扰的路径损耗随距离下降很慢，ASE就会减小，这一点在意料之中。略感意外的是，最优的复用距离也在减小。

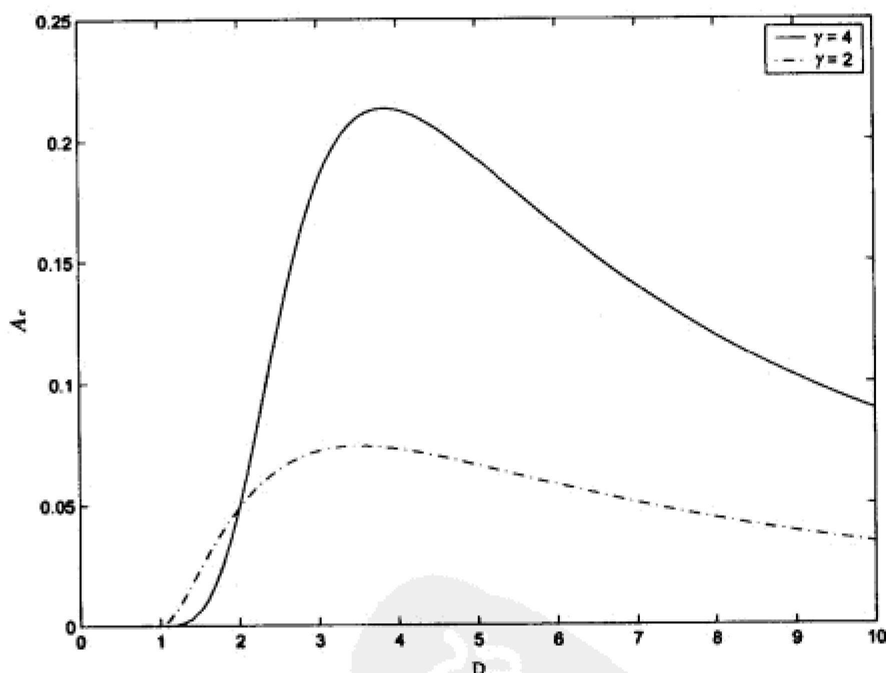


图15-10 AWGN上行链路的面积频谱效率（$\gamma = 2$和4）

现在假设干扰源不在小区边界，它们离期望用户基站的距离是$D-R/2$，则ASE为

$$A_e = \frac{\log_2(1+(D-R/2)^\gamma/6R^2)}{\pi(0.5D)^2} \tag{15-23}$$

图15-11画出了$\gamma = 4$时的ASE，同时还画出了干扰源在小区边界时的ASE。这个结果是预料之中的，干扰源移向它的小区中心时，ASE将增大，同时最佳复用距离也在减小。 527

文献[69]研究了正交信道化的蜂窝系统上行链路的ASE，其中考虑了可变速率传输、以及小区间干扰的最好情况、最坏情况和平均情况。文中还考虑了衰落模型、小区半径、系统负载等的影响。结果表明，在干扰条件为最佳和平均的情况下，最优复用因子是1。

就是说，即使不能靠扩频来抑制干扰，信道也应该在每个小区内复用。此外，ASE随着小区半径按指数的四阶多项式而减小，从而给出了小区缩小和容量增加之间的定量关系。文献[70]在类似框架下研究了下行的ASE。

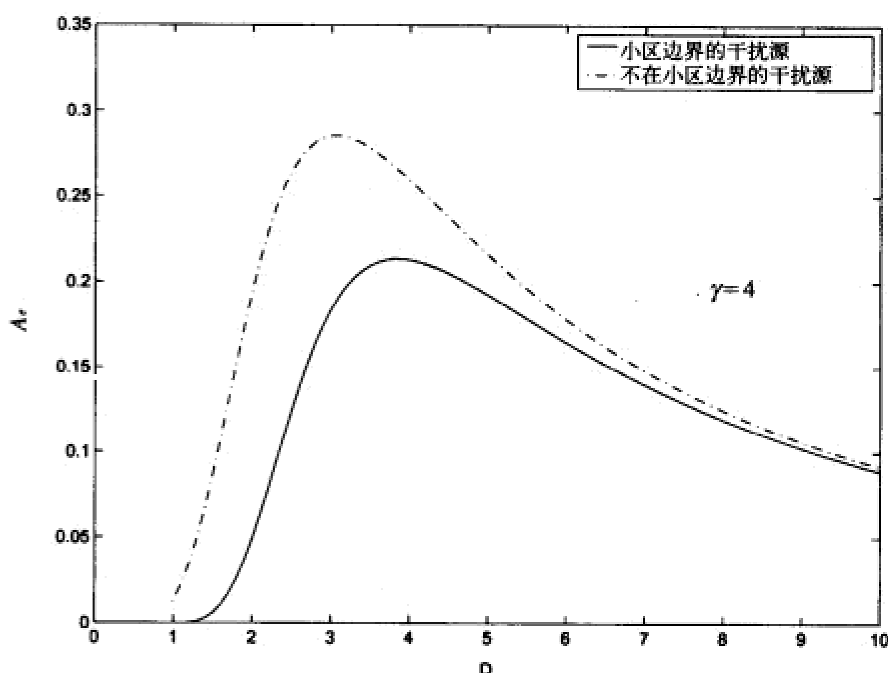


图15-11 干扰源在$D-R/2$和D处时的ASE（$\gamma=4$）

习题

15.1 有一个10km^2的城市，宏蜂窝系统设计将这个城市分为许多1km^2的正方形小区，每个小区能容纳100个用户。计算系统所能容纳的用户总数，以及用户以车速30km/h穿过一个小区所花的时间（用以估计经过多少时间会发生切换）。如果小区缩减到100m^2，并且系统也按比例缩减以保证在这个小小区内仍能容纳100个用户，计算系统所能容纳的用户总数，以及穿越一个小区所需的时间。

15.2 对于15.2节所述的正六边形小区信道分配算法，证明复用距离$D=\sqrt{3}R\sqrt{i^2+j^2+ij}$。

15.3 一蜂窝系统的小区是半径为R = 100m的菱形。假设为了维持要求的SIR，同频小区的中心距离至少必须是$D=600\text{m}$。

(a) 计算所要求的复用因子N和每小区簇中的小区个数。

528

(b) 如果系统的信道总数是450，计算每个小区可以分配到的信道数。

(c) 画出两个相邻的小区簇，按要求的复用距离标出两个小区簇的信道分配情况。

15.4 一蜂窝系统的小区是半径为R = 1km的正六边形。假设为了维持要求的SIR，同频小区的中心距离至少必须是$D=6\text{km}$。

(a) 计算要求的复用因子N和每小区簇中的小区个数。

(b) 如果系统的信道总数是1200，计算每个小区可以分配到的信道数。

(c) 画出两个相邻的小区簇，按要求的复用距离标出两个小区簇的信道分配情况。

15.5　某TDMA蜂窝系统的小区是半径为R = 10m的菱形，复用距离为D = 60m。假设小区内的路径损耗指数为$\gamma_I = 2$，小区间干扰的路径损耗指数为$\gamma_0 = 4$。计算相应的SIR，并与$\gamma = \gamma_I = \gamma_0 = 4$、$\gamma = \gamma_I = \gamma_0 = 2$时的SIR进行比较。解释SIR的大小次序。

15.6　某TDMA蜂窝系统采用正六边形小区，所有信号传播的路径损耗指数都是$\gamma = 2$，系统采用BPSK调制，信道是AWGN信道，要求的误码率为$P_b = 10^{-6}$，系统总带宽为B = 50MHz，每个用户的信号带宽为100kHz。计算最小复用距离和系统的用户容量。

15.7　考虑一个理想功控的CDMA系统，处理增益为G = 100，扩频码参数为$\xi = 1$，比值$\lambda = 1.5$。假设AWGN信道，采用BPSK调制，目标误码率为$P_b = 10^{-6}$。计算无扇区划分时和N = 3扇区化时的系统容量。

15.8　本题考虑话音激活因子的影响，它使干扰数成为一个随机量。CDMA系统的SIR为

$$\mathrm{SIR} = \frac{G}{\sum_{i=1}^{N_c-1}\chi_i + N}$$

其中G是系统的处理增益；χ_i是伯努力分布的小区内干扰，分布概率$\alpha = p(\chi_i = 1)$等于话音激活因子；N代表小区间干扰，假设它是高斯分布，均值为$0.247N_c$、方差为$0.078N_c$。定义中断率为SIR低于目标值SIR_0的概率：$P_{out} = p(\mathrm{SIR}<\mathrm{SIR}_0)$。

(a) 证明 $P_{out} = p\left(\sum_{i=1}^{N_c-1}\chi_i + N > \frac{G}{\mathrm{SIR}_0}\right)$。

(b) 写出P_{out}的表达式。

(c) 利用 (b) 中的表达式计算$N_c = 35$，$\alpha = 0.5$，$\mathrm{SIR}_0 = 5$(7dB)、$G = 150$时的中断率。

(d) 假定N_c足够大，以致于随机变量 $\sum_{i=1}^{N_c-1}\chi_i$ 可以近似为一个高斯随机变量。写出此时 $\sum_{i=1}^{N_c-1}\chi_i + N$ 的概率分布，以及相应的中断率表达式。

(e) 用 (d) 中的表达式计算$N_c = 35$，$\alpha = 0.5$，$\mathrm{SIR}_0 = 5$(7dB)、$G = 150$时的中断率。并与 (c) 的结果进行比较。

15.9　某蜂窝系统有K个用户，每个用户下行最小需要的SIR分别为 $\gamma_1^*,\cdots,\gamma_K^*$，写出满足要求的功控矢量存在的条件。

15.10　考虑半径为R = 1的正六边形小区TDMA上行链路，假定小区内所有用户分配的时间比例相同，发射功率也相同，所有用户都在小区边界上，画出相应的ASE与复用距离D，$0 \leqslant D \leqslant 10$的关系曲线。计算中假设路径损耗指数$\gamma = 2$，干扰仅来自于第一圈干扰小区。干扰源各以1/5的概率出现在下面5个位置：最接近被干扰小区的小区边界；基站和这个最近小区边界的中点；小区中心；离被干扰小区最远的小区边界；基站和这个最远小区边界的中点。并画出干扰源以概率1分别位于最近小区边界和最远小区边界这两种情形下的ASE和最佳复用距离。比较这些曲线的不同。

15.11　考虑一个沿高速公路部署的一维蜂窝系统，小区是边长为$2R$ = 2km的正方形，如图15-12所示。本题主要考虑基站到移动台的下行传输。假设每个小区有两个移动台，位置如图中所示，相对于各自的基站，每个小区内的移动台位置完全一样。每个基站的总发射

529

功率为P_t = 5W，平分给小区内的两个用户。系统总带宽为100kHz，每个接收机的噪声功率谱密度为10^{-16}W/Hz。传播模型为$P_r = P_t K(d_0/d)^3$，其中d_0 = 1m、K = 100。所有干扰可视为加性白高斯噪声，可忽略第一圈干扰小区外的干扰。系统采用频分，每个基站将分给它的带宽平分给小区内的两个用户。

530

(a) 若复用距离$D = 2$（频率隔小区复用），求分配给每个用户的带宽。

(b) 对于快变瑞利衰落，BPSK调制，目标误码率为10^{-3}，求最小复用距离D。

(c) 不计任何衰落和阴影效应，请根据用户速率的香农容量公式，计算频率隔小区复用时（$D = 2$），每个小区的面积频谱效率。

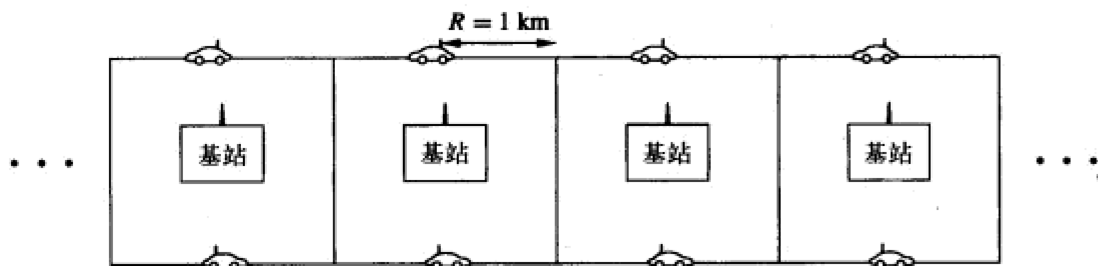


图15-12 正方形小区的一维蜂窝系统（题15.11）

15.12 本题讨论不同系统参数下，蜂窝系统上行链路的每用户的容量。假定上行总带宽为B=100kHz，每个小区内基站接收机的噪声功率谱密度为$N_0/2$，其中$N_0 = 10^{-9}$ W/Hz，并且所有移动台有相同发射功率P。

(a) 对于用户数K = 10，用户发射功率P = 10mW，用式（15-18）画出每用户容量与$\alpha(0 \leqslant \alpha \leqslant 1)$的关系曲线，并解释曲线形状与$\alpha$的关系。

(b) 对于$\alpha = 0.5$，$P = 10$mW，画出每用户容量与用户数$K(1 \leqslant K \leqslant 30)$的关系曲线，并解释曲线形状与$K$的关系。

(c) 对于$\alpha = 0.5$、$K = 10$，画出每用户容量与$P(0 \leqslant P \leqslant 100\text{mW})$的关系曲线，并解释曲线形状与$P$的关系。

参考文献

[1] T. S. Rappaport, *Wireless Communications – Principles and Practice*, 2nd ed., Prentice-Hall, Englewood Cliffs, NJ, 2001.

[2] V. Erceg, A. J. Rustako, and R. S. Roman, "Diffraction around corners and its effects on the microcell coverage area in urban and suburban environments at 900 MHz, 2 GHz, and 4 GHz," *IEEE Trans. Veh. Tech.*, pp. 762–6, August 1994.

[3] A. J. Goldsmith and L. J. Greenstein, "A measurement-based model for predicting coverage areas of urban microcells," *IEEE J. Sel. Areas Commun.*, pp. 1013–23, September 1993.

[4] S. Dehghan and R. Steele, "Small cell city," *IEEE Commun. Mag.*, pp. 52–9, August 1997.

[5] H. Holma and A. Toskala, *WCDMA for UMTS Radio Access for Third Generation Mobile Communications*, 3rd ed., Wiley, New York, 2004.

[6] M. V. Clark, V. Erceg, and L. J. Greenstein, "Reuse efficiency in urban microcellular networks," *IEEE Trans. Veh. Tech.*, pp. 279–88, May 1997.

[7] M. K. Simon, *Probability Distributions Involving Gaussian Random Variables*, Kluwer, Dor-

drecht, 2002.
[8] G. L. Stuber, *Principles of Mobile Communications*, 2nd ed., Kluwer, Dordrecht, 2001.
[9] R. Prasad and A. Kegel, "Effects of Rician faded and log-normal shadowed signals on spectrum efficiency in microcellular radio," *IEEE Trans. Veh. Tech.*, pp. 274–81, August 1993.
[10] R. Kohno, R. Meidan, and L. B. Milstein, "Spread spectrum access methods for wireless communications," *IEEE Commun. Mag.*, pp. 58–67, January 1995.
[11] K. S. Gilhousen, I. M. Jacobs, R. Padovani, A. J. Viterbi, L. A. Weaver, Jr., and C. E. Wheatley III, "On the capacity of a cellular CDMA system," *IEEE Trans. Veh. Tech.*, pp. 303–12, May 1991.
[12] P. T. Brady, "A statistical analysis of on–off patterns in 16 conversations," *Bell System Tech. J.*, pp. 73–91, January 1968.
[13] V. V. Veeravalli and A. Mantravadi, "The coding–spreading trade-off in CDMA systems," *IEEE J. Sel. Areas Commun.*, pp. 396–408, February 2002.
[14] P. Jung, P. W. Baier, and A. Steil, "Advantages of CDMA and spread spectrum techniques over FDMA and TDMA in cellular mobile radio applications," *IEEE Trans. Veh. Tech.*, pp. 357–64, August 1993.
531
[15] T. S. Rappaport and L. B. Milstein, "Effects of radio propagation path loss on DS-CDMA cellular frequency reuse efficiency for the reverse channel," *IEEE Trans. Veh. Tech.*, pp. 231–42, August 1992.
[16] B. Gundmundson, J. Sköld, and J. K. Ugland, "A comparison of CDMA and TDMA systems," *Proc. IEEE Veh. Tech. Conf.*, pp. 732–5, May 1992.
[17] J. Winters, "Smart antennas for wireless systems," *IEEE Pers. Commun. Mag.*, pp. 23–7, February 1998.
[18] S. Verdú, *Multiuser Detection*, Cambridge University Press, 1998.
[19] B. M. Zaidel, S. Shamai, and S. Verdú, "Multicell uplink spectral efficiency of coded DS-CDMA with random signatures," *IEEE J. Sel. Areas Commun.*, pp. 1556–69, August 2001.
[20] J. G. Andrews, "Interference cancellation for cellular systems: A contemporary overview," *IEEE Wireless Commun. Mag.*, pp. 19–29, April 2005.
[21] S. Shamai and B. M. Zaidel, "Enhancing the cellular downlink capacity via co-processing at the transmitting end," *Proc. IEEE Veh. Tech. Conf.*, pp. 1745–9, May 2001.
[22] H. Viswanathan, S. Venkatesan, and H. Huang, "Downlink capacity evaluation of cellular networks with known-interference cancellation," *IEEE J. Sel. Areas Commun.*, pp. 802–11, June 2003.
[23] S. A. Jafar, G. J. Foschini, and A. J. Goldsmith, "PhantomNet: Exploring optimal multicellular multiple antenna systems," *J. Appl. Signal Proc.* (EURASIP), pp. 591–605, May 2004.
[24] M. Costa, "Writing on dirty paper," *IEEE Trans. Inform. Theory*, pp. 439–41, May 1983.
[25] E. H. Choi, W. Choi, and J. Andrews, "Throughput of the 1x EV-DO system with various scheduling algorithms," *Proc. Internat. Sympos. Spread Spec. Tech. Appl.*, pp. 359–63, August 2004.
[26] D. Wu and R. Negi, "Downlink scheduling in a cellular network for quality-of-service assurance," *IEEE Trans. Veh. Tech.*, pp. 1547–57, September 2004.
[27] L. Xu, X. Shen, and J. W. Mark, "Dynamic fair scheduling with QoS constraints in multimedia wideband CDMA cellular networks," *IEEE Trans. Wireless Commun.*, pp. 60–73, January 2004.
[28] X. Qiu, L. Chang, Z. Kostic, T. M. Willis, N. Mehta, L. G. Greenstein, K. Chawla, J. F. Whitehead, and J. Chuang, "Some performance results for the downlink shared channel in WCDMA," *Proc. IEEE Internat. Conf. Commun.*, pp. 376–80, April 2002.
[29] A. C. Varsou and H. V. Poor, "HOLPRO: A new rate scheduling algorithm for the downlink of CDMA networks," *Proc. IEEE Veh. Tech. Conf.*, pp. 948–54, September 2000.
[30] E. Villier, P. Legg, and S. Barrett, "Packet data transmissions in a W-CDMA network – Examples of uplink scheduling and performance," *Proc. IEEE Veh. Tech. Conf.*, pp. 2449–53, May 2000.

PDG

[31] L. Qian and K. Kumaran, "Uplink scheduling in CDMA packet-data systems," *Proc. IEEE Infocom Conf.*, pp. 292–300, March 2003.

[32] L. Qian and K. Kumaran, "Scheduling on uplink of CDMA packet data network with successive interference cancellation," *Proc. IEEE Wireless Commun. Network Conf.*, pp. 1645–50, March 2003.

[33] H. Boche and M. Wiczanowski, "Queueing theoretic optimal scheduling for multiple input multiple output multiple access channel," *Proc. IEEE Internat. Sympos. Signal Proc. Inform. Tech.*, pp. 576–9, December 2003.

[34] K.-N. Lau, "Analytical framework for multiuser uplink MIMO space-time scheduling design with convex utility functions," *IEEE Trans. Wireless Commun.*, pp. 1832–43, September 2004.

[35] L. F. Chang, X. Qiu, K. Chawla, and C. Jian, "Providing differentiated services in EGPRS through radio resource management," *Proc. IEEE Internat. Conf. Commun.*, pp. 2296–2301, June 2001.

[36] I. Katzela and M. Naghshineh, "Channel assignment schemes for cellular mobile telecommunication systems – A comprehensive survey," *IEEE Pers. Commun. Mag.*, pp. 10–31, June 1996.

[37] D. C. Cox, "Wireless network access for personal communications," *IEEE Commun. Mag.*, pp. 96–115, December 1992.

532

[38] R. J. McEliece and K. N. Sivarajan, "Performance limits for channelized cellular telephone systems," *IEEE Trans. Inform. Theory*, pp. 21–4, January 1994.

[39] D. Everitt and D. Manfield, "Performance analysis of cellular mobile communication systems with dynamic channel assignment," *IEEE J. Sel. Areas Commun.*, pp. 1172–81, October 1989.

[40] J. Zander and H. Eriksson, "Asymptotic bounds on the performance of a class of dynamic channel assignment algorithms," *IEEE J. Sel. Areas Commun.*, pp. 926–33, August 1993.

[41] M. R. Garey and D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*, Freeman, New York, 1979.

[42] D. Kunz, "Channel assignment for cellular radio using neural networks," *IEEE Trans. Veh. Tech.*, pp. 188–93, February 1991.

[43] R. Mathar and J. Mattfeldt, "Channel assignment in cellular radio networks," *IEEE Trans. Veh. Tech.*, pp. 647–56, November 1993.

[44] A. Lozano and D. C. Cox, "Distributed dynamic channel assignment in TDMA mobile communication systems," *IEEE Trans. Veh. Tech.*, pp. 1397–1406, November 2002.

[45] L. Chen, U. Yoshida, H. Murata, and S. Hirose, "Dynamic timeslot allocation algorithms suitable for asymmetric traffic in multimedia TDMA/TDD cellular radio," *Proc. IEEE Veh. Tech. Conf.*, pp. 1424–8, May 1998.

[46] Y. Hara, T. Nabetani, and S. Hara, "Performance evaluation of cellular SDMA/TDMA systems with variable bit rate multimedia traffic," *Proc. IEEE Veh. Tech. Conf.*, pp. 1731–4, October 2001.

[47] A. C. Kam, T. Minn, and K.-Y. Siu, "Supporting rate guarantee and fair access for bursty data traffic in W-CDMA," *IEEE J. Sel. Areas Commun.*, pp. 2121–30, November 2001.

[48] U. C. Kozat, I. Koutsopoulos, and L. Tassiulas, "Dynamic code assignment and spreading gain adaptation in synchronous CDMA wireless networks," *Proc. Internat. Sympos. Spread Spec. Tech. Appl.*, pp. 593–7, November 2002.

[49] D. Ayyagari and A. Ephremides, "Cellular multicode CDMA capacity for integrated (voice and data) services," *IEEE J. Sel. Areas Commun.*, pp. 928–38, May 1999.

[50] E. Biglieri, G. Caire, and G. Taricco, "CDMA system design through asymptotic analysis," *IEEE Trans. Commun.*, pp. 1882–96, November 2000.

[51] J. Zander, "Performance of optimum transmitter power control in cellular radio systems," *IEEE Trans. Veh. Tech.*, pp. 57–62, February 1992.

[52] G. J. Foschini and Z. Miljanic, "A simple distributed autonomous power control algorithm and its convergence," *IEEE Trans. Veh. Tech.*, pp. 641–6, November 1993.

[53] S. A. Grandhi, R. Vijayan, and D. J. Goodman, "Distributed power control in cellular radio sys-

tems," *IEEE Trans. Commun.*, pp. 226–8, February–April 1994.

[54] S. A. Grandhi, R. D. Yates, and D. J. Goodman, "Resource allocation for cellular radio systems," *IEEE Trans. Veh. Tech.*, pp. 581–7, August 1997.

[55] A. Lozano and D. C. Cox, "Integrated dynamic channel assignment and power control in TDMA mobile wireless communication systems," *IEEE J. Sel. Areas Commun.*, pp. 2031–40, November 1999.

[56] R. Veronesi, V. Tralli, J. Zander, and M. Zorzi, "Distributed dynamic resource allocation with power shaping for multicell SDMA packet access networks," *Proc. IEEE Wireless Commun. Network Conf.*, pp. 2515–20, March 2004.

[57] F. Berggren, S.-L. Kim, R. Jantti, and J. Zander, "Joint power control and intracell scheduling of DS-CDMA nonreal time data," *IEEE J. Sel. Areas Commun.*, pp. 1860–70, October 2001.

[58] H. C. Akin and K. M. Wasserman, "Resource allocation and scheduling in uplink for multimedia CDMA wireless systems," *IEEE/Sarnoff Sympos. Adv. Wired Wireless Commun.*, pp. 185–8, April 2004.

[59] T. H. Hu and M. M. K. Liu, "A new power control function for multirate DS-CDMA systems," *IEEE Trans. Commun.* pp. 896–904, June 1999.

[60] S.-J. Oh, D. Zhang, and K. M. Wasserman, "Optimal resource allocation in multiservice CDMA networks," *IEEE Trans. Wireless Commun.*, pp. 811–21, July 2003. 533

[61] D. Ayyagari and A. Ephremides, "Optimal admission control in cellular DS-CDMA systems with multimedia traffic," *IEEE Trans. Wireless Commun.*, pp. 195–202, January 2003.

[62] S. V. Hanly and P. Whiting, "Information theory and the design of multi-receiver networks," *Proc. Internat. Sympos. Spread Spec. Tech. Appl.*, pp. 103–6, November 1992.

[63] A. Wyner, "Shannon-theoretic approach to a Gaussian cellular," *IEEE Trans. Inform. Theory*, pp. 1713–27, November 1994.

[64] T. Cover and J. Thomas, *Elements of Information Theory*, Wiley, New York, 1991.

[65] E. C. van der Meulen, "Some reflections on the interference channel," in R. E. Blahut, D. J. Costello, and T. Mittelholzer, Eds., *Communications and Cryptography: Two Sides of One Tapestry*, pp. 409–21, Kluwer, Boston, 1994.

[66] M. H. M. Costa and A. A. El Gamal, "The capacity region of the discrete memoryless interference channel with strong interference," *IEEE Trans. Inform. Theory*, pp. 710–11, September 1987.

[67] S. Shamai and A. D. Wyner, "Information-theoretic considerations for symmetric, cellular, multiple-access fading channels: Part I," *IEEE Trans. Inform. Theory*, pp. 1877–94, November 1997.

[68] S. Shamai and A. D. Wyner, "Information-theoretic considerations for symmetric, cellular, multiple-access fading channels: Part II," *IEEE Trans. Inform. Theory*, pp. 1895–1911, November 1997.

[69] M.-S. Alouini and A. J. Goldsmith, "Area spectral efficiency of cellular mobile radio systems," *IEEE Trans. Veh. Tech.*, pp. 1047–66, July 1999.

[70] M. F. Tariz and A. Nix, "Area spectral efficiency of a channel adaptive cellular mobile radio system in a correlated shadowed environment," *Proc. IEEE Veh. Tech. Conf.*, pp. 1075–9, May 1998. 534

第16章 无线自组织网络

无线自组织网络（ad hoc wireless network）是一些无线移动节点的集合，它们无需借助事先建立的设施即可自行构建成一个网络，如图16-1所示。在没有网络架构的情况下，这些移动节点靠自己来处理必要的控制和网络功能，一般是通过分布式控制算法实现这一目的。多跳路由通过中间结点向目的节点转发分组包，它能够改善网络的吞吐量和功率效率。韦氏字典（Merriam Webster）对“ad hoc”这个词给出的两个重要定义是：“formed or used for specific or immediate problems（为特定或当前问题而形成或使用的）”和“fashioned from whatever is immediately available（用立即可得的东西制成的）”。这些定义准确地反映了无线自组织网络的两个主要优点：它能自动适应具体的应用；只要有可用的网络节点，就可以组网。无线自组织网络还有其他吸引人之处：不需要投资、建设和维护基础设施；可以快速部署和重组；分布式特性、节点冗余性和不存在单点故障问题使它有很好的健壮性和抗毁性。这些特征对于军事应用尤为重要，许多关于无线自组织网络的突破性研究都受到过美国国防部高级研究计划署（DARPA）和美国海军的支持[1-6]。这些早期的工作研究确定了许多无线自组织网络的基本设计原则。然而在过去几十年里，尽管无线通信，特别是无线自组织网络总体上取得了很大的进步，但是人们对无线自组织网络的最佳设计、性能、网络的能力等问题仍然没有很好的理解，至少与其他无线网络比较是这样的。

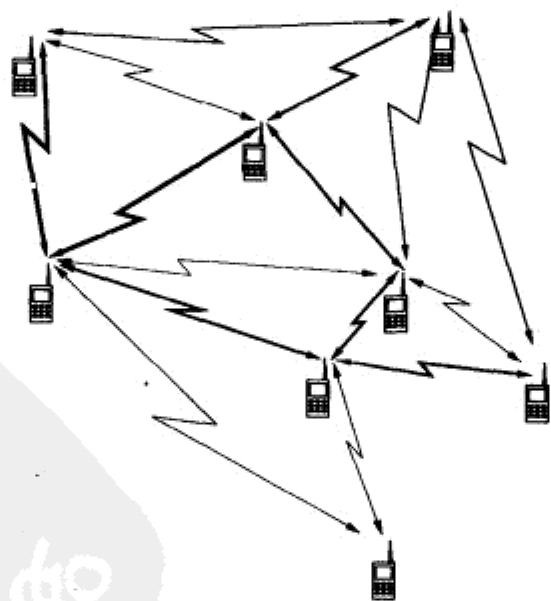


图16-1 自组织网络

因为应用驱动许多设计要求，所以本章首先总述无线自组织网络的主要应用，然后介绍网络的基本设计原则及其面临的挑战。接着介绍协议分层的概念、层间的相互作用以及跨层设计的优点。本章还将简要介绍网络的基本容量极限和规律。最后讨论能量受限的无线自组织网络所特有的一些关键问题。

16.1 应用

本节介绍无线自组织网络最主要的一些应用。自组织和无架构虽然会导致性能上的损失但仍然对许多应用非常有吸引力。无需基础架构对低成本的商用系统非常适合，因为不需要网络建设和运营的大规模投资，网络的部署成本与网络的发展程度成比例。无基础架构这个特点对军事系统也很重要。军事通信网一般处在偏远地区，而且在需要它的时候必须能迅速建成。无线自组织网络的其他优点还包括网络易重构、维护成本低。

不过，无线自组织网络所固有的多跳路由、分布式控制会造成性能损失，必须权衡考虑其优点和损失。

我们主要考虑如下几个应用：数据网络、家庭网络、设备网络、传感器网络和分布式控制系统。这并不是无线自组织网络应用的全部。实际上，无线自组织网络的成功来自于其充分的灵活性。这里存在一个两难的网络设计问题。如果网络有最大的灵活性，能支持许多应用（通用型网络），就难以按具体的应用要求去裁剪网络。这将使一些应用的性能变差，特别是高速率和时延严格限制的应用。另一方面，如果网络是针对一些特定应用设计的，设计者就必须能提前预测杀手级应用是什么——这是有风险的。理想的无线自组织网络应该足够灵活以支持多种不同的应用，同时又能根据运营中的应用随时调整性能。自适应跨层设计有这种灵活性，并能根据节点的限制条件来裁剪协议设计。 536

16.1.1 数据网络

自组织无线数据网络主要支持笔记本电脑、掌上电脑、个人数字助理（PDA）及其他信息工具间的数据交换。这些数据网络一般按其覆盖区域可分为三类：局域网（LAN）、城域网（MAN）和广域网（WAN）。虽然低成本、高性能的架构式无线局域网已经非常流行，但无线自组织数据网络有它自己的优势。首先，借助多跳路由，一个接入点就能把整个区域内的所有节点都连接到有线骨干网，这就减小了成本和安装要求。其次，有时候节点通过接入点或基站进行通信是很低效的，如两个相邻的PDA，它们能够直接交换信息，而不需要从别处转发。

无线城域网覆盖范围大，一般需要多跳路由，跳数通常也较多。这种网络的难点在于如何以低成本的方式通过多跳网络传输高速数据，其中每跳的链路质量可能不同，并且还是时变的。没有集中的网络控制、用户可能有高移动性，都使这个问题更为复杂。例如美国国防部高级研究计划局（DARPA）在全球移动信息系统（Global mobile information systems，GLOMO）等军用项目上投入了大量的时间和金钱，以期建立一个能支持多媒体的高速自组织无线城域网，但所取得的成果非常有限[7-8]。无线自组织城域网也已经渗透到了商业领域，最好的例子就是Metricom[9]。虽然Metricom在几个主要的城市提供了较高的数据速率，但一直没有出现显著的需求，不得不走向破产。

无线广域网适用于不能部署网络架构，或者部署成本太高的大区域。例如在不能部署网络基础架构的偏僻地区，我们可以洒一个传感器网络。此外，军事、救灾等需要快速构建和拆除网络的情形，只能依靠自组织网络。 537

16.1.2 家庭网络

家庭网络用于支持居室内或附近设备之间的通信，包括台式电脑、笔记本电脑、个人数字助理、无绳电话、智能家电、监控系统、消费电子产品和娱乐设备。这种网络能实现“智能房间”，它能感知到有没有人及人的移动，再相应调节照明和采暖；能实现“感知家庭”，它将传感器和计算机组成网络，能照顾老弱病残的生活。家庭网络能使视频监控或传感监测系统有综合解读数据的智能，可在异常情况下向主人、警察或者消防

队发出告警；能使智能家电互连并能连到互联网，从而可以进行远程控制、软件升级和规划维护；能使娱乐系统中随便一个电视机或音响系统都能连接到VCR、机顶盒或个人计算机[10~13]。

设计这样的网络将面临一些挑战性的问题。最主要的挑战之一是不同的家庭网络应用将导致变化的服务质量（QoS）要求。这里的QoS指特定应用的一些要求，一般是数据速率和时延限制，家庭娱乐系统的要求可能很严。其他主要的挑战包括成本和标准化问题，家庭网络中的设备都必须遵循相同的网络标准。家庭网络中不同的设备在功耗限制上差别很大，有些用固定电源，基本没有功耗受限的问题，而有些是用电池供电，或许还不能充电。因此，家庭网络设计中最大的挑战之一就是让功率不受限的设备尽可能多地承担通信和网络负担，使得整个网络中不论功耗限制若何，所有节点的网络要求都能满足。

16.1.3 设备网络

设备网络提供设备之间的短距离无线连接，其目的是用无线连接来替代很麻烦的线缆连接，使蜂窝电话、调制解调器、耳机、PDA、计算机、打印机、投影机、网络接入点以及其他设备都不再需要缆线和相应的接头。引导这方面技术的是一些低成本低功率的无线组网技术，如蓝牙[14~15]、紫蜂[16]和超宽带（UWB）[17]。商用电子设备集成了这样的无线技术就可以互连。一些常见的应用有：蜂窝电话的无线耳机、无线USB连接器、无线RS232连接器、无线PCMCIA卡、无线机顶盒等。

16.1.4 传感器网络

无线传感器网络由许多小节点组成，这些节点具有传感、计算和网络功能，它也体现了这三种重要技术的聚合。无论是民用还是军用，传感器网络都有巨大的潜力。军事任务要求传感器或其他信息搜集装置能安置在目标附近，因而它们潜在面临很大的危险，因此所用的技术必须要有很高的冗余性或者很强的抗毁性，并且尽量不需要人工干预。对于这种要求，一种办法是用大量被动传感器组成一个电磁、光、化学、生物的传感器阵列，它可以识别和追踪目标，也可以用作各类进攻的一线监测。这种网络也可用于无人智能车辆的运动，如用光传感器网络提供网络导航、指引车辆绕开障碍物进入防御或进攻位置。一些工业应用中的设计考虑也和这些军事应用类似，比如用传感器阵列来遥感核电站、矿山和其他工业场所。

538

传感器网络在家用方面可以实现电表、煤气表和水表的远程无线读表。在家中大量使用这种简便的计量表，我们就能监测一些高耗电或耗气的家电，如空调、热水器，并相应进行使用规划。这样的计量和通信设备其实就是给电源插头增加一个简单附件。这样我们就可以想象，某个人坐在家用电脑前就能看到家中的各种耗电信息。同样可以监测和控制电视的使用和电视内容。烟雾监测也是家用的一个重要例子，不光是监测各个房屋，还要相互通信以能追踪火势的蔓延情况，在消防员赶到之前，这些信息以及房屋的设计图早已传送到消防队。类似的传感器阵列还可用于探测煤气或其他有毒气体的泄

漏和蔓延情况。

传感器阵列在重大事故中也有巨大的应用潜力。比如在房屋倒塌的营救行动中，传感器阵列可以迅速部署在事故现场，它可以跟踪热量、天然气和有毒物质，可以用声学传感器和定位技术发现被困的幸存者。甚至有可能通过这样的传感器阵列来防止灾难的发生，例如为了预警桥梁、人行道和阳台等的坍塌，可预先在其中嵌入应力传感器和移位传感器。把大量低成本、低功耗的传感器在浇注水泥之前置入建筑物，就能跟踪材料的疲劳损伤。这些传感器必须能自行设置并且要耐用，能和建筑物的寿命相适应。

很多传感器用的电池不能充电，电池寿命问题可以通过超小型的能量收集无线电台来克服。这方面研究预期的目标是：无线电台的体积小于1cm³、重量小于100g、功耗小于100μW[18]。需要的功耗如此低，从环境中吸取能量就够用了，这就是能量收集(energy harvesting)技术。有了这样的无线电台，我们可以把传感器大量部署在建筑物、家庭、甚至是人体内，从而能开拓出许多新的应用。

16.1.5 分布式控制系统

无线自组织网络能提供无线信道把远程站点、传感器和执行器连接在一起，从而实现分布式控制应用。和有线的分布式系统相比，它能够组织协调无人控制的移动单元，能大大减小维护和重构的成本。无线自组织网络可用于自动化高速公路系统中的车辆协调控制、工业制造以及其他工业流程中的远程控制、军事中无人飞行器的协调控制等。 539

现有的分布式控制已经有很好的性能，并对模型参数的不确定性具有健壮性。但这些设计都基于闭环性能，并假设了中心化结构、同步时钟系统和固定拓扑结构。它要求传感器和执行器的信号能以很小的固定时延传送到控制器。无线自组织网络不能保证数据速率、时延和丢包特性，其时延一般具有随机性、数据包也可能丢失。不幸的是，分布式控制大多对通信中的差错不具健壮性，很小的随机时延也可能是灾难性的[19-20]。因此，必须重新设计分布式控制，以使它能对无线网络中的随机延迟和丢包特性有健壮性[21]。理想情况下，可以将无线自组织网络和控制器进行联合设计，以尽可能提高端到端的性能。

16.2 设计原则和挑战

无线自组织网络最基本的特性是没有基础架构，大部分的设计原则和挑战也都来源于此。这一点很容易通过和主流无线网络进行比较来看清楚，如蜂窝系统或无线局域网。蜂窝系统将目标地理区域划分为多个小区，小区内的移动台与小区中心的基站通信，基站连到有线骨干网。移动台之间没有直接的通信，所有的通信都是经过基站的单跳路由。基站和骨干网络完成所有网络功能，包括鉴权、呼叫路由和切换。大部分无线局域网也有一个类似的、中央控制的单跳路由结构：移动节点直接和接入点通信，接入点连到骨干互联网，接入节点完成所有网络和控制功能[1]。与此相反，无线自组织网络中有点对

1. 802.11无线局域网标准包括自组织网络的兼容性，但是标准的这部分很少被用到。

点的直接通信，网络和控制功能分布在各个节点上，路由可能会借助中间节点转发。

无线自组织网络也能以永久或动态的方式形成架构或者节点等级。例如，许多无线自组织网络利用部分节点形成一个骨干架构以解决可靠性、规模和容量的问题[22]。若有一个骨干节点离开了网络，骨干架构可能要重组。也可以选定一些节点来作为邻近节点的基站[14]。因此，无线自组织网络可以创建一个网络结构来改善性能，但这样的结构不在网络的基本设计要求内。

540

没有典型结构对有线网络来说是平常的事情。多数的城域网和广域网都是自组织的结构，包括互联网也是这样。不过在无线自组织网络中，无线电波的广播特性是有线网络所没有的。如果发射功率足够大，任何一个节点都能直接把信号发送给其他任何一个节点。发射功率固定时，两个通信节点间的链路的SINR一般将随距离的增加而减小，同时还与环境的传播特性和干扰有关。此外，SINR还因信号和干扰中的衰落而随时间随机变化。链路的SINR决定了它的通信性能，包括数据速率、分组差错率或误比特率。一般不使用SINR极低的链路，因而整个网络是部分连通的，如图16-1所示。链路的连接性并不是有连或不连这样的二元值，节点能用自适应调制来适应SINR，节点也能用功控来改变它。图16-1中不同的线宽表示不同的SINR。虽然理论上任何一个节点都能直接发送数据到其他任何一个节点，距离很大的两个节点实际上不大可能直接通信，短距离的链路也有可能性能很差，或者会对其他链路造成较大的干扰。网络的连接性也随节点进入和离开网络而变化。新节点进入时，可以通过控制已有节点的发射功率来控制网络的连接性[23]。

链路的许多参数，如功率、数据速率等都是可以变的，于是就有了这种连接的灵活性，它对路由有很大的影响。如果单跳路由的SINR高过最小的门限值，节点就会直接发包到目的地。单跳路由的SINR有可能很差，而且单跳路由也会对周围的节点造成干扰。因此，在大多数的无线自组织网络中，数据包是从源节点出发，经过许多中间转发节点后才到达目的节点的。由于路径损耗使接收功率随距离指数下降，所以中继转发能大大降低端到端传输所需要的总发射功率（源节点和所有中继节点的发射功率之和）。借助中继节点的多跳路由是无线自组织网络的一个重要特征，正是这个特征使地理上分散的节点可以通信、使网络有可扩展性、使网络成为分布控制。然而，相比蜂窝系统和无线局域网的单跳信道，在多跳无线信道上支持高数据速率和低延迟有更大的挑战。这是无线自组织网络支持高数据速率低延迟业务（如视频）时遇到的主要困难之一。

有大量节点的无线自组织网络必须要有可扩展性。可扩展性的关键在于分布式网络控制算法，这些算法根据局部的条件来调整局部的性能。没有集中的信息和控制资源后，协议只依赖于局部信息，因此能适应网络的增长。无线自组织网络中，协议可扩展性方面的研究主要集中在自组织[24-25]、分布式路由[26]、移动性管理[22]和安全性[27]等方面。分布式协议经常要花费许多能量于本地处理和报文交换中，文献[28]针对安全性协议对此进行了详细的分析。于是就有了本地处理和发送信息给中心处理之间怎么折中的问题。这个折中在传感器网络中非常明显，在这里，相近的节点的数据有相关性，它们又一起协调路由来传输这个数据。无线自组织网络可扩展性方面的实验工作大多集中在较小的

网络上，少于100个节点。而许多应用，特别是传感器网络，能有成百上千，甚至更多的节点。现有的无线网络协议是否具有这样的可扩展性还不清楚。

541

能量限制也是无线自组织网络设计中的一个重大挑战[29]。这些限制主要出现在依靠不能充电的电池供电的无线节点上，如传感器。硬性的能量限制对网络设计的考虑有很大的影响。首先，没有了数据速率的概念，因为电池耗尽前每个节点所能传输的比特数量是有限的。比特周期和能量消耗间也有折中的问题，缓慢地发送的比特可以节约发射能量，但会增加电路消耗的能量[30]。待机会消耗许多能量，所以节约能量要用休眠模式，但节点休眠会使网络控制和路由问题变得复杂。实际上，能量约束几乎能影响到所有的网络协议，优化能耗需要考虑网络设计的每个方面。

16.3 协议层

协议分层是网络设计中常用的一种抽象概念。分层使网络协议的设计有了模块性，便于实施和标准化。然而分层的理念不很适合无线自组织网络，这里许多协议的设计问题是交杂在一起的。本节介绍用于无线自组织无线的协议分层，以及不同协议层间的相互作用，后者引出了跨层设计的问题。

人们已经发展出了一种数据网络协议分层框架的国际标准，称为开放系统互联（Open System Interconnection, OSI）模型。OSI模型将网络要求的功能分为七层：应用层、表示层、会话层、传输层、网络层、数据链路控制层和物理层。每一层负责不同的任务，层间用固定的接口交换控制信息和数据。OSI模型的基本思想是，任一层的协议都能与其他层的协议互操作，而不用考虑协议实施的细节。例如应用层不必考虑数据在网络中的路由，也不必考虑在给定的链路上用了什么调制什么编码。所有层的协议集叫做网络的协议栈。关于OSI模型及每一层的功能请参考文献[31]的1.3节。

互联网推动了协议分层的实际应用，它的协议是围绕着传输层的传输控制协议（Transport Control Protocol，TCP）和网络层的互联网协议（Internet Protocol，IP）建立的。大部分网络中的OSI分层模型已经被五层模型所取代，这个模型也称为TCP/IP模型，它由TCP和IP协议的主要功能定义。五层模型包含应用层、传输层、网络层、接入层和物理层。图16-2示出了这五层，并说明了它们在无线自组织网络中的主要功能，下面还会详细介绍。注意功率控制同时位于物理层和接入层上，它也是网络层资源分配的一部分[32]。因此功率控制跨越了协议栈中的多个层。大部分无线自组织网络的路由不用IP协议，因为无线网络的路由与互联网有很大的不同。IP协议中的寻址和子网也不适合无线自组织无线网络。传输协议也不一定要用TCP。五层模型是无线网络中协议层模块设计的常用抽象。

542

协议设计的分层原则在互联网这样的有线网络中运作得很好，在那里，物理层的数据速率高达每秒数千兆比特，而且很少丢包。但即使在这种情况下，分层也使它难以支持有硬时延限制的高数据速率应用，如视频，甚至话音。无线网络的物理层可能数据速率非常低、分组和比特错误率非常高。在这种情况下，协议分层大大降低了效率，同时也阻碍了利用协议层之间的相互作用来改善性能。跨层设计同时考虑协议栈中的多个层，

或者是联合设计，或者是在层间交换信息。跨层设计比严格的分层方法有很大的优越性。下面我们介绍五层模型和它们在无线网络中的功能，然后讨论跨层设计的基本原理及其性能优势。

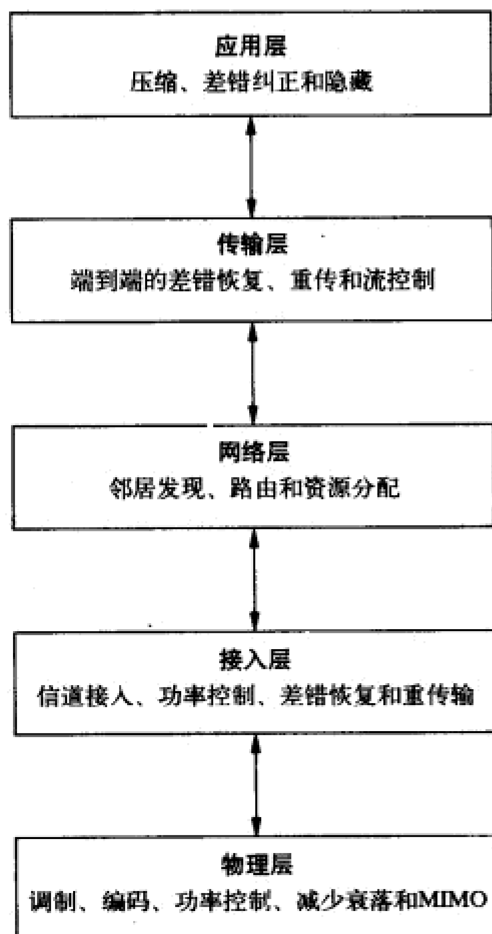


图16-2 网络协议设计的五层模型

16.3.1 物理层设计

物理层主要涉及点对点无线链路上的比特传输，因此也叫*链路层*（link layer）。本书第5章～第13章全面讨论了物理层中的各种设计问题，包括调制、编码、分集、自适应技术、MIMO、均衡、多载波调制和扩频。链路是无线自组织网络的一部分，物理层的不同设计对其上的协议层也有影响。实际上，无线自组织网络的物理层中，大部分设计因素都会以某种方式影响高层协议。下面我们给出几个例子，通过物理层设计时对分组差错速率、多天线和功率控制等因素的不同考虑，来说明物理层和高层的相互作用。

543

大部分的无线自组织网络是按14.3节所述的分组方式传输的。物理层的设计连同信道和干扰条件决定了链路的*分组差错率*（packet error rate，PER）。许多接入层协议将重传接收出错的数据分组，因此物理层设计的PER影响着接入层的重传请求。类似的，10.8节指出，多天线设计时可在复用/分集/方向性之间进行折中：可用多天线来增加链路的数据速率，可用来实现分集以降低衰落信道中的平均BER，也可以通过方向性来减轻衰落和对其他链路的干扰。分集能减小PER，从而能降低重传次数。复用能提高链路速率，从而能减小拥塞和时延，有利于经由该路的所有多跳路由。方向性能减小对其他链路的干扰，从而能改善它们的性能。因此，无线自组织网络中如何最佳地使用多天线这个问题明显超出了物理层的范围，它同时涉及物理层、接入层、网络层和传输层。

一个节点在物理层的发射功率也影响到协议栈中的多个层。加大发射功率将减小PER，从而会影响接入层的重传请求。实际上，只要发送功率足够大，网络中的任意两个节点都能直接通信，所以功率控制着连接性。一个节点过高的发射功率将严重干扰其他节点，降低这些节点的性能，并有可能使某些链路中断连接。链路性能表现为SINR，因此所有节点的发射功率将影响到网络中所有链路的性能。更宽一些说，一个节点的发射功率以及相适应的自适应调制和编码决定了这个节点的“本地邻域”，即单跳路由能

到达的节点集合，因此也决定了接入、路由和其他高层协议操作的范围。因此，网络中所有节点发射功率的优化要考虑到它所影响的所有协议层。这就是提出跨层间设计的主要原因。

16.3.2 接入层设计

接入层控制不同用户对频谱的共享，并要确保一共享频谱上的用户的数据分组能够成功接收。信号维在用户间的分配是通过多址接入或随机接入来实现的，关于这些接入技术的细节见14.2节和14.3节。多址接入以正交或非正交的信道化方法将信号维分割成一些专用的信道，最常用的方法是TDMA、FDMA和CDMA。接入层还必须有一些控制功能来给用户分配信道，并在系统不能容纳用户时拒绝其接入。在随机接入中，信道以动态的方式分配给激活用户，在多跳网络中这些协议还必须要解决隐藏终端和暴露终端的问题。常见的随机接入是不同形式的ALOHA、CSMA和调度（见14.3节）。这些随机接入方法包含有信道分配和拒绝接入的功能。

544

前一节的讨论说明，单个节点的发射功率会影响到其他所有节点。因此，对所有节点的功率控制是接入层功能的一部分。功率控制的主要作用是要保证网络中所有的链路都能满足SINR目标值。从下面的详细讨论中我们将看到，这一点通常是不可能。14.4节和15.5.3节分别针对多址接入和蜂窝系统所讨论的，以满足SINR目标值为目的的功控算法可以扩展到自组织网络。假设一个无线自组织网络有K个节点，这些节点的发射接和接收机之间有N条链路[1]。链路k的SINR为

$$\gamma_k = \frac{g_{kk}P_k}{n_k + \rho\sum_{j\neq k} g_{kj}P_j} \quad k, j \in \{1,2,\cdots,N\} \tag{16-1}$$

其中$g_{kj} > 0$是第j条链路的发射机到第k条的链路接收机之间的信道功率增益，P_k是第k条链路发射机的功率，n_k是第k条链路接收机的噪声功率，ρ是信号处理引入的干扰减小因子（处理增益为G的CDMA系统$\rho \approx 1/G$，TDMA中$\rho=1$）。假定第k条链路要求的SINR是γ_k^*，这个值例如可由连通性要求和数据速率要求来决定。于是，所有链路的SINR约束可以表示为矩阵形式$(\boldsymbol{I}-\boldsymbol{F})\boldsymbol{P} \geqslant \boldsymbol{u}$, $\boldsymbol{P} > 0$，其中$\boldsymbol{P} = (P_1, P_2, \cdots, P_N)^T$是$N$条链路的发射功率列向量，$\boldsymbol{u}$是将噪声乘以SINR约束值，除以信道增益后得到的列向量：

$$\boldsymbol{u} = \left(\frac{\gamma_1^* n_1}{g_{11}}, \frac{\gamma_2^* n_2}{g_{22}}, \cdots, \frac{\gamma_N^* n_N}{g_{NN}}\right)^T \tag{16-2}$$

矩阵$\boldsymbol{F}$的元素为：

$$F_{kj} = \begin{cases} 0 & k = j \\ \gamma_k^* g_{kj}\rho/g_{kk} & k \neq j \end{cases} \tag{16-3}$$

其中$k, j \in \{1,2,\cdots,N\}$。如同14.4节和15.5.3节所讨论的控制问题，如果$\boldsymbol{F}$的Perron-

1. 前面已经提到，所有节点都能和其他节点通信，于是若网络有K个节点，则链路有$K(K-1)$条。我们假设只用其中的$N \leqslant K(K-1)$条，因此只需考虑这N条链路的SINR。

16.3.3 网络层设计

网络层负责建立和维护网络中端到端的连接。这就要求网络是完全可连接的[1]，即网络中的任何一个节点都可以直接或经由中间节点与任何其他节点通信。无线自组织网络中网络层的主要功能是邻居发现、路由和动态资源分配。邻居发现是节点首次进入网络时发现其邻居的过程。路由是网络层的另一个关键功能，它决定从源节点到目的节点的包传输的路径。经由中间节点的路由一般是中继转发，也可以考虑其他方法来更好地利用多用户分集。动态资源分配决定功率、带宽等资源在网络中的分配。总的来说，资源分配发生在协议栈的多个层上，所有跨层设计是有必要的。

1. 邻居发现和拓扑控制

邻居发现是节点随机分布的网络初始化所要完成的最先的几个步骤之一。对于节点来说，邻居发现就是要确定出哪些节点是可以直接通信的，这样的直接通信功率不超过最大级别，链路性能（指数据速率和相关的BER）不低于最低级别。允许的发送功率越高，可发现的邻居节点也越多。不同协议层上的不同寻址方式将会使邻居发现复杂化，这个问题必须要解决。

邻居发现过程先用某个初始发射功率来探测邻居节点。如果这个功率不能建立起$N\geqslant 1$个邻居连接，就提高功率、再重复探测。一直到建立了N个连接，或者已达到最大发射功率P_{max}为止。参数N根据网络要求的最小连接性设置，P_{max}根据每个节点的功率限制和网络设计来设置。如果N或P_{max}过小的话，会出现不能连通的情形，此时网络成为若干个节点簇，簇内能相互通信，但不能到达别的簇。图16-3示出了这种情形，图中以节点为圆心的虚线环代表了它可以建立连接的邻居域。如果N与P_{max}过大，许多节点的发送功率对于全连接（网络一般都是全连接的）来说就显得过于浪费，这会增加干扰。对于全连接的网络，可以采用类似于式（16-4）的分布式功控算法来以最小的发射功率满足所有链路的SINR要求。另外，也可以用功率控制来构造出所需要的网络拓扑结构[8]。 547

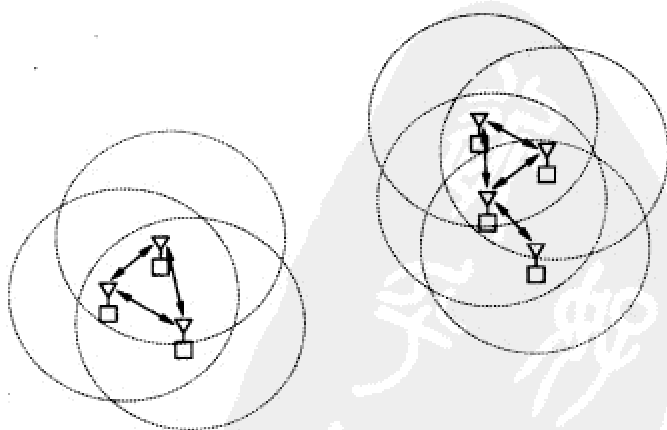


图16-3 不能连通的网络

1. 注意图论中全连接的含义是，图中每个节点都有一条边连到每个其他的节点。

Frobenius（最大模）特征值小于1，那么存在矢量$P > 0$（即对所有k有$P_k > 0$）使所有链路都能达到要求的SINR值，且Pareto最优解为$P^* = (I-F)^{-1}u$，也即P^*是所有满足SINR要求的解中发射功率最小的[33-35]。此外，下面的分布式迭代算法能收敛到最优解P^*[34-35]。第k条链路在第$i+1$时刻的发射功率更新为

545

$$P_k(i+1) = \frac{\gamma_k^*}{\gamma_k(i)} P_k(i) \tag{16-4}$$

这是一个简单的分布式功控算法，第k条链路的发射机只要求已知此链路的SINR。如果这个SINR低于目标值，就提高发射功率，反之就降低发射功率。如此简单的分布式算法能收敛到全局最优的功率控制的确是相当出色的。但而当信道不是静态时，这个算法不能确定可满足SINR约束条件，此时也很难找出一种能满足要求的分布式功控算法[36]。当信道时变时，式（16-4）给出的算法会使链路的SINR出现大幅度的波动性。更为重要的是，即使对于静态的链路，也往往不能满足所有的SINR要求，其原因是因为网络中干扰源数目众多，而且信道增益的也很大。当SINR目标不能满足时，分布式功控算法将会发散，最终所有节点都将以最大功率发射，但仍不能满足SINR要求。这显然是不希望有的运行状态，特别对于能量受限的节点。

新用户进入系统时，接纳控制可以保护已有链路的SINR，使它仍在目标值以上。文献[37]扩展了前述静态链路的分布式功控算法，使其包括了分布式接纳控制。该方案让激活链路的SINR稍微高出目标值一点。新用户先以低功率尝试接入系统，此时已有链路所留出的SINR缓冲量就可以保护这些激活链路的SINR不会低于最小目标值。为了达到SINR目标值，新用户再逐渐增加功率。如果接纳这个新用户不会影响已有链路的SINR限制，那么在新用户递增功率的过程中，分布式算法最终将收敛到一个新的P^*，它能使包括新链路在内的所有链路都能达到SINR要求。如果系统不能接纳新用户，递增功率也不能使其达到所要求的SINR，新用户最终被迫离开系统。不过在时变信道下，分布式控制很难有一个优化的接纳准则[36]。将这些思想和传输调度相结合[38-39]，可以提高功率效率、减小干扰。功率控制算法可以通过考虑时延约束得到修正[23, 40]。不过，时延约束与多跳路由中分组包的整个路由有关，为了满足端到端的时延约束，功率控制还应该与网络层协议协同工作。

接入层也负责错误分组的重传，通常称为*自动重传*（automatic repeat request, ARQ）协议。数据分组中包含了检错编码，接收端用它来判决接收到的数据分组中是否存在不可纠正的比特错误，这样的数据分组一般会被丢弃，接收端通过反馈信道通知发射机重传。接入层也可以不丢弃错误的分组，而是将它保存下来，以便同重传的分组一起形成某种分集合并，这样能大大提高重传的正确接收率。也可以不把原始数据分组整体重传，而是发送一些附加的编码比特，使接收端有更强的纠错能力，这种技术称为*递增冗余*（incremental redundancy），发送端只要发送了足够多的冗余比特就能纠正原始数据分组中的错误比特。与简单重传相比，上述的分集和递增冗余能显著提高吞吐量[41]。

546

为了使网络能够全连接，每个节点具体需要的邻居数与具体的网络配置、信道特性等有关。对于随机分布的静止节点，只考虑路径损耗时一般需要有6~8个邻居节点[2, 31]。更一般假设下，有关所需邻居数的分析见[42-44]。当节点的移动性增加时，衰落将使链路的增益剧烈变化。此时如果不增加功率来补偿瞬时衰落，就难以保证在任何时间上都能保持全连接。如果数据对时延不敏感的话，衰落倒可以通过网络分集来改善网络连接[45]。网络密度减小时，连接性会出现问题[43, 46-48]。物理层的调适能力，比如速率、功率和编码方式等，也对连接性有很大影响，即使在SINR很低的链路上，调整这些参数后也能实现通信。

2. 路由

无线自组织网络中的路由协议面临很大的挑战，特别节点移动的情况下，连接性的快速变化要求能够动态地重构路由。几十年来，围绕无线自组织网络的路由协议，人们已经进行了广泛深入的研究。很难对这些工作进行简单的分类，我们将关注三类主要的路由协议[49]：洪泛、主动路由（集中式、源驱动或者分布式）和被动路由。

洪泛（flooding）把数据包广播给接收范围内的所有节点，这些节点又继续广播这个数据包，一直这样进行下去，直到这个数据包到达它的最终目的地。洪泛对动态变化的网络拓扑结构有很强的健壮性，而且路由开销也很少。在高速移动的网络中，洪泛或许是唯一的选择。洪泛的缺点是同一个数据包的多个副本在网络中传输，浪费了传输节点的带宽和电池功率。因此，除了那些非常小的网络外，洪泛并不是一个实际可行的方法。

548

与洪泛相反的思路是*集中式*（centralized）路由计算。它让所有的节点把它所测定的信道状态信息和网络拓扑信息发送给一个中心，这个中心负责计算网络中每个节点的路由表，再把这些表传给节点。计算最佳路由的准则与希望优化的目标有关，一般包括最小平均时延、最少跳数和最小网络拥塞[50]。这些目标一般可以映射为路由上每跳的一个代价。于是，源和目的地之间的最小代价路由可以用一些经典的优化技术得到，如Bellman-Ford算法或者Dijkstra算法[51]。虽然集中式路由计算能给出特定优化条件下最高效的路由，但它不能适应信道状态或网络拓扑的快速变化，并且定期搜集各个节点的本地信息以及发布路由信息也会导致较大的开销。集中式路由算法也和洪泛算法一样，只适用于非常小的网络。

集中式路由算法的一个变型就是*源驱动*（source-driven）路由。在这里，每个节点有整个网络的连接信息，节点用这些信息自己计算到达某个目的地的最佳路由。源驱动路由也要定期搜集网络的连接信息，同样会产生较大的开销。集中式路由和源驱动路都可以和分级路由相结合，就是把节点归到不同等级的簇中，在每个簇内来执行路由。

分布式（distributed）路由计算是无线自组织网络中应用最普遍的路由算法。在这个协议中，节点将它的连接信息发送给相邻节点，再根据这些本地信息来计算路由，具体就是确定出路由的下一跳是哪里。分布式路由有许多优点。首先，本地节点之间交换路由信息的开销非常小。其次，这种算法能快速适应链路特性和连接性的变化。其缺点是用本地信息得到的全局路由一般是次优的，并且常常会出现路由环路，即数

据分组会在某些中间节点间上循环，而不能到达目的节点。**目的序列距离矢量**（destination sequenced distance vector，DSDV）协议在路由表中加入了序号，从而可以避免路由环路[52]。

集中式和分布式路由都有一个按一定准则优化的路由表。按最小跳数优化就叫距**离矢量路由**（distance vector routing），按每跳的一般性代价函数优化就叫做**链路状态路由**（link state routing）。由于网络的动态性，不论是集中式还是分布式的路由优化算法，路由表都要定期更新。另一种方法是被动（按需）路由，它只在源节点有数据要发送时才创建路由。这就节省维护路由表的开销，因为当前不用的路由不需要维护。在这一方法中，源节点有数据要发送时，它将发起路由发现过程。这个过程测定出是否有一个或多个到达目的节点的路由。这一条或多条路由一直要被保持到发完数据。被动路由的优点是只需相对较少开销即可获得高效的全局高效的路由，因为这些路由不必一直维护。被动路由的缺点是会引起较大的初始延迟，因为只有当有数据要发送时，才会发起路由发现过程，要等到路由发现过程结束才能传输数据。最常见的按需路由协议是**自组织按需距离矢量**（ad hoc on-demand distance vector，AODV）协议[53]和**动态源路由**（dynamic source routing，DSR）协议[54]。被动和主动路可以结合为一种混合形式，叫**区域路由协议**（zone routing protocal，ZRP），它能减小被动路由的延迟，并能减少主动路由的开销[55]。

549

移动性对路由协议有非常大的影响，它会使已经建立的路由不再存在。高移动性会严重降低主动路由的性能，因为路由表很快就变得过期了，保持实时更新需要增加大量的开销。高移动性时，洪泛算法在维持路由方面非常有效，但在网络效率方面代价也非常大。洪泛的一种修正叫多径路由，它不用增加大量开销就能非常有效。多径路由把一个数据分组通过若干个不同的端到端路由传输，这些路由上传输的分组不大可能都丢失或者都有太大的时延，故而这个分组很可能会在某一条路径上以最小时延到达目的地[56]。这种技术在拓扑结构动态变化的场合有很好的性能。

路由协议是基于网络拓扑的，路由建立只能建立在有合理质量的链路上。然而如前所述，两个节点之间连接性的含义是有弹性的，它依赖于链路的SINR，而物理层的设计又决定了可靠传输数据所需要的SINR。接入层决定了链路间的干扰，因而也对连接性也有重要影响。因此，物理层、接入层和网络层之间有密切的相互作用[57]。文献[38]研究了这种相互作用，发现CSMA/CA结合了使用低SINR链路的路由协议时，网络的吞吐量将会显著减小。文献[38]还指出，如果在任意的源目的节点对之间只维持单条路由，则对于网络总的吞吐量来说这只是次优的方案。若在任意源目的节点对之间建立多条路由，用这些路由复用传输，则能改变对其他端到端路由造成的干扰，利用这种方法可以增加网络的吞吐量。

路由算法也可以按照高层协议的要求来优化，特别是应用层的时延和数据速率要求。这样的算法叫做QoS路由。QoS路由的目标是在网络中找出一个端到端时延和数据速率能满足业务要求的路由，其实例和性能见文献[58~60]。

550

大部分路由协议在中继节点上采用译码转发。中继节点对收到的数据包进行译码，

并通过纠错去除错误，如果发现有错并且不能纠正时，就要求重传。另一种方法是放大转发，中继节点只是简单地重传收到的分组包，不进行检错和纠错。这种方式简化了中继设计，减小了中继节点的能耗和时延。不过放大转发在无线系统中效果不太好，因为无线链路不可靠、常常有差错，这些问题会逐跳累积。相对于这两种方法的另一选择是合作分集（cooperative diversity），它通过空间上不同的多个节点共同转发同一用户的数据分组，从而形成分集[61~62]。这个想法最初在文献[63~64]中提出，其中多个收发信机重发自己检测出的其他信号，形成了一个带空间分集的重复码。在此基础了，又进一步出现了更复杂的合作分集及合作编码技术[65~66]。最后，网络编码通过聚合多条路由接收来的数据来增加网络容量[67~69]。网络编码主要用于有线网络的多播传输中，也能用于无线系统[70]。

3. 资源分配和流控制

路由协议确定出数据分组从源节点到目的节点应该走的路由。当路由的优化准则是最小拥塞或者延迟时，路由功能就和传输层的流控制功能有重叠。如果路由算法在某个链路上发送的数据太多的话，这条链路就会出现拥塞，路由算法就必须要另找一个新的路由。此外，链路时延是链路的数据速率或容量的函数，链路的容量越高，就能有更多的数据以最小的时延通过这个链路。链路的容量又取决于分配给它的资源，即发射功率和带宽，因此我们可以看到，路由、资源分配和流控制是相互关联的。

忽略处理时延和传播时延，节点i到节点j的链路时延的经典度量为（见文献[31]的5.4节）：

$$D_{ij} = \frac{f_{ij}}{C_{ij} - f_{ij}} \tag{16-5}$$

其中f_{ij}是分配给这个链路的业务流量，C_{ij}是链路的容量（链路能支持的最大流量）。这个公式来自排队论，它对实际应用是一个很好的度量。链路上的流量越接近于最大数据速率时，就越有可能发生拥塞和延迟。链路的另一个度量是其利用率，即

$$D_{ij} = \frac{f_{ij}}{C_{ij}} \tag{16-6}$$

根据文献[31]的5.4节，这个度量的特性和时延度量（16-5）相似，也是流量和容量的拟凸函数[1]，使得路由算法可以利用高效的凸优化方法。如果流经每条链路的流量是固定的，路由算法就可以根据式（16-5）的时延度量或者式（16-6）的利用率度量来算出每跳的代价，进而可以找出最小代价的路由。这两个度量的差别在于，随着流量逐渐接近链路容量，时延将趋于无限大，而链路利用率是趋于1。因此当流量接近于链路容量时，式（16-5）的代价大大高于式（16-6）。新的路由建立之后，它将会改变这个路由上的链路流量，不过一般改变不大，因为单个节点对总业务量的贡献很小。但对于中小型网络，视频等数据量大的业务可能会引起严重的拥塞。

551

1. 如果一个函数在低于任意给定的门限值的部分都是上凸的，这个函数就是拟凸（quasi-convex）函数。

链路度量（16-5）和（16-6）都假定链路容量C_{ij}是固定的，但这个容量实际上是链路SINR和所分配带宽的函数。通过动态分配给拥赛的链路多分一些功率或带宽可以增加其容量并减小时延，但这会使其他链路的资源变少，使它们的容量降低。链路容量的改变反过来又造成链路度量的改变，而链路度量又是计算最优路由的依据，这样就影响到整个网络的性能。因此，网络的性能同时与路由、流控制和资源分配有关。 552

假定代价和容量是凸函数（或拟凸函数），那么流控制、路由和资源分配的联合优化能够表示为流量和通信变量的凸优化问题，可用内点凸优化方法来求解。利用这种方法，文献[71~73]研究了联合路由算法和资源分配算法时，TDMA和CDMA无线网络中最小化功率、最大化链路利用率或最大化流量利用率的问题。使用迭代优化的类似想法见文献[74]。在这种情况下的最大吞吐量可能造成系统资源的高度不公平分配[75]，但可以进行修正以纳入公平性约束[76]。

16.3.4 传输层设计

传输层提供端到端的差错恢复、重传、重排序和流控制功能。虽然单条链路也有差错检测和重传，但这些机制不是十分完善的。传输层通过这样两点提供了更强的保护：(i) 监测端到端的分组差错或丢失，(ii) 如果发现分组错误或丢失，则要求源节点重传。另外，多路由、延迟和拥塞、分组丢失引起的重传都可能造成分组到达的次序不是原来的次序。传输层把分组的次序排好以后再送给应用层。

通过把应用层的数据流分配到不同的路由，传输层也能提供网络的流控制功能。传输层的TCP协议在无线网络中表现不佳，因为它假定分组丢失都是拥塞引起的，因此会启动拥塞控制。有线网络中拥塞是分组丢失的主要原因，所以TCP协议工作得很好，这也是它被用于互联网传输层的原因。但无线网络中分组丢失的主要原因是信道的不理想因素和节点的移动性。此时调用拥塞控制将使效率严重降低（见文献[49]的11.5节)。通过把链路失效信息反馈给传输层，目前在改善无线网络TCP性能方面已经取得了一些有限的进展。

总的来说，如16.3.3节所述，无线网络的流量控制与资源分配和路由算法有复杂的关系。在无线网络中，这种关联性要比有线网络紧密得多。特别是，有线网络的链路容量是固定的，而无线链路的容量取决于链路间的干扰。在某个链路上分配一个业务流会引起对其他链路的干扰，进而影响这些链路的容量和时延。这种关联性使我们很难把流控制、资源分配和路由功能按照网络层和传输层方式分开，这就提出了对这两个层进行跨层设计的问题。 553

16.3.5 应用层设计

应用层产生在网络上发送的数据，并处理从网络上接收到的数据。这一层提供对业务数据的压缩、差错纠正和隐藏功能。数据业务的压缩必须是无损的，视频、话音或图像业务重建时能容忍一定的损失，可以采用有损压缩。压缩级别越高，加给网络的数据速率负担也越小。但高度压缩的数据去掉了大部分的冗余，所以对差错更为敏感。数据

业务不能容忍任何损失，端到端传输中损坏或丢失的分组必须被重传，这会引起很大的时延。语音、视频和图像业务能够容忍一些差错，采用差错隐藏或自适应重放等技术能减小这些差错对感观质量的影响[77~78]。因此，应用层的设计需要在数据速率及健壮性之间折中：速率越高，带给网络的负担也越重，但同时对网络性能的健壮性也越强。

应用层也能通过*多描述编码*（multiple description coding，MDC）来提供分集形式[79~80]。MDC是一种压缩方式，它产生数据的多重描述。原始数据能以一定的损失从任何一种描述中重建，描述越多，重建的性能就越好。当源数据的多个描述通过网络发送时，丢失、延迟或者损坏一些描述不会对整体性能造成太大的影响。因此，MDC是应用层针对不可靠的网络所设计的一种分集。此外，MDC结合多路径路由能通过描述和传输这些描述的路径来形成跨层分集[81]。设计的折中在于，对于给定的数据速率，MDC的失真比单纯的压缩技术大，可以看成是性能和分集的折中。通过牺牲一些性能，我们能换取对网络不确定性的健壮性。

许多应用要求保证端到端的数据速率和时延，统称为QoS。今天的互联网，即使在高速高质量固定通信链路下，也不能保障端到端的QoS。在无线自组织网络中，链路的容量低、差错率高而且是时变的，用户在移动，网络拓扑在动态变化，在这种情况下要保证这样的QoS是不现实的。应当采用能适应网络的时变QoS参数。虽然在前几节所述的物理层、接入层和网络层中的自适应能提供尽可能好的QoS，但这个QoS会因为信道条件、网络拓扑和用户需求的改变而随时间变化。因此业务必须能适应于网络所能提供的QoS。也可以通过约定，降低不重要用户的QoS，以此来改善高优先级的用户的QoS。

举个简单的例子，网络可以根据低层协议的能力得出一个速率和时延的折中曲线[82]，应用层则需决定在曲线上的哪一点上运行。某些应用（如某些数据业务）能容忍更高的延迟，但不能接受较低的整体速率。另一些应用（如分布式控制）可能对延迟非常敏感，但能容忍较低的速率（如粗量化的传感数据）。语音、视频等允许一定失真的应用可以通过降低对误码的健壮性来换取更高的速据速率。能量约束也能引出网络性能和网络寿命的折中。因此，网络设计的折中曲线一般是多维的，包括速率、延迟、健壮性和寿命。这些折中还随网络中用户数量的变化和环境的变化而变化。

16.4　跨层设计

没有中央控制、没有骨干架构、无线链路的特性各不相同，这一切使无线自组织网络很难支持大量出现的应用，特别是要求高数据速率和严格时延的应用。在无线和有线网络的分层设计中，协议栈的每一层都独立于其他层的设计和操作，这种方法总的效果不是很好，特别对于严格的性能要求。分层方法排除了前几节所讨论的联合优化所能带来的好处。此外，针对孤立的单层所设计的好的协议，往往会在层间产生消极的影响，会显著降低端到端的性能，同时也使网络对于动态变化和干扰极度脆弱。因此，无线自组织网络中严格的性能要求只能通过跨层设计来满足。这样的设计需要对层间的依赖关系进行分析和利用，并进行联合优化。很明显，跨层设计要求层间能交换信息，要求每一层的协议对这些信息有适应性，并要求每层都有分集机制以保障健壮性。

554

虽然无线和有线网络都可以采用跨层设计，但无线信道的特征使这样的设计框架对无线自组织网络来说有独特的挑战和机遇。节点间的链路即能用于节点通信，也能使它们互相干扰，这个链路可以通过自适应协议，如自适应调制和编码、自适应空时频信号处理和自适应功控来控制。高层协议（接入和路由）依赖于节点的连接性和干扰，因此高层协议可以利用物理层的自适应来获得更好的性能。同时，某些链路会出现严重的拥塞或衰落，高层协议可以通过自适应路由来跳过这些链路，从而能降低这些链路引起的延迟和瓶颈。最高层可以用端到端路由的吞吐量和延迟信息来改变压缩率，或者在多路由上发送MDC数据。这样，高层协议就能自适应于低层的状态。

协议栈中每一层的自适应要和该层所适应的时间变化相应。链路SINR的变化非常快，快衰落时能达到微秒级变化。网络拓扑的变化要慢得多，为秒级。而用户的流量根据业务之不同，在数十到数百秒范围内变化。网络变化的不同时间尺度表明每一层协议应该首先考虑补偿本层的时间变化。如果本层自适应不能补偿这个时间变化，就应该向高层协议发送信息，以期在更大的范围内解决这一问题。例如，假定端到端路由上一给定链路的SINR比较低。这个连接性信息被传送到协议栈高层（即要求网络层重新路由，或要求应用层采用低速率压缩）的时候，链路的SINR可能已经改变。因此，每一层的协议应该适应本层的变化。如果本层的适应性不足以补偿这种性能衰减，上一层的性能将会因此而恶化。上一层的自适应应能纠正或至少能减轻下一层自适应未能解决的问题。仍考虑低SINR链路的例子。物理层能够快速而准确地测量出链路的SINR。因而物理层协议能够通过增加发射功率或降低编码率来响应低SINR，这将能校正如由多径平衰落引起的连接性变化。但如果链路质量差是物理层难以纠正的原因造成的，如移动台进入隧道，那么让上一层来处理这个问题就更好，比如推迟分组的传输，直到移动节点离开隧道。类似地，如果网络中的节点高速移动，那么链路特性和网络拓扑将快速变化，将高移动节点通知给网络层可以改变路由策略，使其从单播改成期望用户总体方向上的广播。最终，如果网络不能提供业务所要求的QoS，那么业务必须能适应于它可以得到的QoS，否则只能承受性能的恶化。这种适应网络的整合方法，即在给定高层自适应的情况下，本层如何响应本地的变化的方法，构成了自适应跨层协议设计。 555

分集是另一个跨层设计可以利用的机制。分集通常用于物理层以提高对衰落的健壮性。但分集的基本思想可以扩展到网络协议栈的每一层。接入层的合作分集通过多个空间分布的节点来辅助分组的转发，能提高对任意链路上的分组损坏的健壮性。网络层的分集体现为多路径路由，它通过多条路由来发送单个分组。这种分集引入了网络层上分集和吞吐量的折中，类似于物理层中的MIMO（见10.5节）。通过多条路由传输的分组不太可能在所有路由上都丢失或者都有严重时延。这样，网络分集就降低了丢包率和平均时延。不过这个分组使用了本可用于发送其他分组的网络资源，于是降低了网络的吞吐量。应用层分集来自于描述业务数据的MDC，只要收到一个描述，就能重建源数据（虽然和全部描述比有些失真）。跨所有层的分集，并结合自适应跨层设计将能保证无线自组织网络的可靠性和良好性能，尽管还存在一些内在的挑战。

前几节已经讨论了应用层以下的跨层设计问题。包括应用层在内的跨层间设计是一

个要求多学科专业知识的困难挑战，目前这方面的工作还很少。另外，跨层设计也存在一些潜在的缺陷，包括复杂度增加和结构的灵活性减小[83]。即便如此，跨层设计潜在的性能增益还是很显著的，文献[84~86]中针对视频和图像传输而进行的无线自组织网络跨层设计已经表明了这一点。

跨层设计对能量受限的网络尤为重要。在这里，每个节点只有有限的能量，必须要针对协议栈的所有层进行优化。能量约束带来了跨层设计独有的挑战和机遇。16.6.4节将讨论其中的一些设计问题。

16.5 网络的容量极限

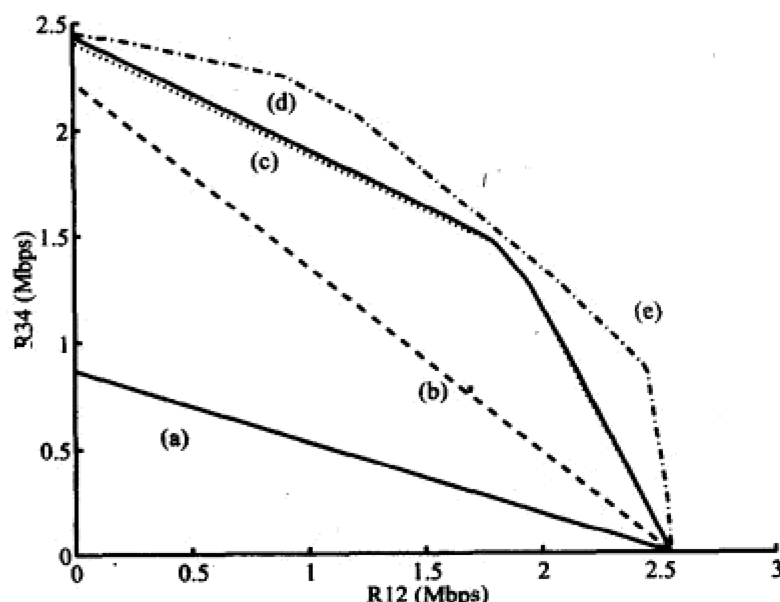
无线自组织网络的基本容量极限，即所有节点间可达的最大速率集合，是信息论中的一个挑战性问题。对于有K个节点的网络，每个节点能和其他$K-1$个节点通信，所以节点间发送独立的信息时容量域的维数为$K(K-1)$。同时发送独立信息和公共信息（多播）的容量域的维数要大得多。即使对于少量的节点和简单的信道配置（如一般的中继和干扰信道），无线自组织无线网络的容量问题也还没有解决[87]。虽然文献[87]的定理14.10.1指出，任意两个不相交的节点集合的速率和的上界是相应的互信息，但对于无线自组织网络的容量域，把这个公式简化为可处理的表达式是一个非常复杂的问题。

556

虽然求解完整的容量域似乎不太可能，降低求解目标也能对问题的实质有所了解。一个里程碑性的结果是Gupta和Kumar在文献[88]中得到了节点数K充分大时，网络吞吐量随K变化的标度律。他们发现，随K的增大，每个节点的吞吐量逐渐下降，下降速率在$1/\sqrt{K\log K}\sim1/\sqrt{K}$之间。换句话说，虽然网络总的吞吐量（等于每节点速率乘以K）以$\sqrt{K/\log K}$和$\sqrt{K}$之间的速率在增长，但每个节点的速率趋于零。这一令人惊奇的结果表明，即使采用最优的路由和调度，大规模无线自组织无线网络中每节点的速率仍趋于零。原因在于中间节的大量资源花在了转发分组到其他节点上，只剩很少的资源能用来发送他们自己的数据。这个结果在一定程度上说是偏于保守的，因为它假定节点随机选择它们的目的节点，但在许多网络中，节点间的通信大部分发生在本地。文献[89]进一步扩展了这些结果，证明如果移动节点能在其物理位置移动到目标节点附近时发送信息，那么这种移动性事实上能把每节点的速率提高到一个常量，即移动性能增加网络容量。这种增加的原因在于移动性使网络不断变化，利用这种变化就能改善每个用户的速率。当然这会产生显著的时延。对于充分大的的固定和移动网络，文献[90~91]研究了吞吐量和时延之间的折中。类似的思想也被用于有限大小的网络和中继网络[43,92]。

除了标度律外，还可以在次优的传输策略下计算可达速率域。文献[45]用这种方法得到了时分策略下各种可能的速率矩阵构成的可达速率域。速率矩阵表示所有源—目的节点对在任意时刻能支持的速率。在不同时隙上取速率矩阵的凸组合，可以得到时分策略下所有源—目的节点对之间的可达速率。速率矩阵取决于当时传输的节点、所有链路的SINR以及传输策略。传输策略越有效，给定矩阵中的数据速率就越大，时分方式下可用的矩阵数也越多。文献[45]中考虑的策略包括可变速率传输、单跳或多跳路由、功率控制和串行干扰抵消。研究框架也可以包括移动性和衰落的影响。图16-4示出了一个

五节点网络的二维剖面，节点在一个正方区域内随机分布，假设节点间的信号传播服从路径损耗指数为γ=4的简化路径损耗模型。这个图是20维速率域的二维剖面，它说明两对节点间的可达速率，这两对节点是节点1～2，节点3～4，此时网络中所有其他的节点可能会用来转发这两对上的分组，但不产生自己的独立数据。图中还假设了基于链路SINR的可变速率传输。图中画出了不同条件下的可达速率域，包括单跳或多跳路由、空间复用、功率控制和串行干扰抵消。可用看到，多跳路由、空间复用和干扰抵消能显著增加容量。功率控制对容量的增加不明显，这是因已经采用了自适应调制，再增加功控不会有太多的不同，至少对这个特定的网络配置来说是这样的。 557



(a) 单跳路由，无空间复用，(b) 多跳路由，无空间复用，(c) 多跳路由，空间复用，

(d) 在 (c) 中增加两级功控，(e) 在 (c) 中增加串行干扰消除

图16-4 五节点网络的容量域沿平面R_{ij}=0, $\{ij\} \neq \{12\}, \{34\}, i \neq j$的剖面

文献[93~97]研究了不同形式的合作分集下的网络容量域。由于一般无线自组织网络的容量域未知，所以合作情形下的容量主要用基于可达速率域的下界，或者基于速率和互信息的上界来反映。结果表明合作能显著增加容量，但合作带来的增益以及能带来这种增益的合作技术都高度依赖于网络的拓扑结构和可用的信道信息。

16.6 能量受限网络

许多无线自组织网络的节点靠寿命有限的电池来供电，因此，网络设计必须要考虑到能量约束的影响。采用可充电电池时应当尽量节约能量，以增加充电间隔。嵌入墙中的、或洒在偏僻地区的传感器网络是无法充电的，它们必须能工作许多年，而供电只能依靠电池或从周围环境中收集的能量。µ-AMPS和PicoRadio项目的目标就是为这样的应用研

558 发低耗电无线设备，其功耗低于100μW，并能利用能量收集来延长寿命[18, 98~99]。

能量约束将影响硬件操作、发射功率和相关的信号处理。对于给定的BER目标和噪声信道，将信号能量扩散到所有可能的时间和带宽维上可以使发送每个比特所需要的能量最低[100]。不过发射功率不是耗电的唯一因素。与分组发送和接收相关的信号处理，甚至待机时的硬件运行的耗电也是不可忽略的[24, 101~102]。这就带来了各层协议设计时的能量折中问题。在物理层上，许多能减小发射功率的技术都需要大量的信号处理。一般认为信号处理需要的能量很小，并随硬件技术的进展而持续降低[24, 103]。但文献[101~102]中的结果表明这些能量耗也是很显著的。这说明在能量受限系统的设计中，必须要使协议栈各层信号处理所耗费的能量最小，也必须要减少用于网络控制的报文传递，因为这些操会显著消耗发射能量和接收能量。节点的休眠模式也要做类似的优化，休眠虽然节约了待机的能量，但接入和路由的复杂化将会增加其他协议层的能耗。节点的硬件和操作系统也可以进行优化以节约能耗[101, 104]。事实上，能量约束影响着协议栈的每一层，也使跨层设计变得更为重要[23, 105~107]。本节介绍能量受限的无线自组织网络设计中的一些关键考虑。

16.6.1 调制和编码

对调制和编码方式的选择一般需要基于发射功率、数据速率、BER和复杂性进行折中。短距离通信中，模拟及数字的信号处理电路所耗费的电能能够和发射功率可比。此时应按总耗电来进行设计，包括发射耗电和电路的消耗。电路的能耗很难建立起一个适当的模型，它与具体使用的硬件密切相关[108]，使得电路能耗和发射能耗之间折中很难一般化。但这样的折中确切存在，特别对于发射能量非常低的短距离通信。

电路消耗的能量随发射信息的时间增加，把发射时间降到最小，再转到休模式能显著节约能量。文献[98]表明，对于短距离应用可以采用多进制调制，发射完成后立即关闭所有电路，这样做能比二进制调制节约能量。文献[109]针对MQAM调制对这种方法进行了分析，并提供了能使总能耗最小的最优策略。文献[30]针对MQAM和MFSK调制，AWGN和瑞利衰落信道，进一步把这种方法扩展到联合优化调制带宽、发射时间和星座大小。结果表明，根据距离来优化发射时间能显著减小能耗：距离大时，发射功率起支配作用，宜选择发射时间长的小星座，距离短时则相反。其结果是，短距离时，MQAM的能量效率比MFSK稍好，因为它的发射时间更短；长距离时，非线性放大的能

559 量效率使MFSK更好。

能量约束也会改变编码自有的折中问题。对于给定的BER目标，编码一般可以降低每比特所需要的发射能量，但这是以编译码处理能量为代价的。此外，许多编码（如分组码和卷积码）编码后的码字长度长于原信息长度，有时也称此为带宽扩展。在给定BER目标下，虽然编码后的码字所需要的总发射能量或许低于未编码时，但码字的发送时间变长，这样电路消耗的能量也变多。编码与调制联合的技术，如格码调制不会扩展带宽，因而也没有带宽扩展引起的能量损失，但需要考虑其编译码处理需要的能量，才能得知它是否可以净省能量。文献[30]研究了能量约束对编码调制的MQAM和MFSK的

影响。结果表明，格码调制的MQAM对能量的节约出现在几乎所有有意义的发射距离上（对于文中所考虑的硬件参数，有意义的距离是1m以上）。而编码MFSK一般带宽效率不高，编码对MFSK的益处只出现在中等发射距离处（对文中条件，指大于30m）。

16.6.2 MIMO及合作MIMO

多入多出技术可以通过复用或分集增益极大地增强无线系统的性能。对于给定的每比特发射能量，复用增益能提供更高的数据速率，分集增益能使衰落信道中的BER更低。然而，MIMO系统在电路方面要比单天线系统耗费更多的能量，因为每根天线都要有单独的电路，而且MIMO相关的信号处理有可能非常复杂。因此，目前尚不清楚在能量约束的条件下，MIMO技术是否还有性能增益。文献[110]研究了这个问题，发现对于大部分有意义的发射距离，按距离优化了星座大小的MIMO确实比单天线系统更省能量。原因在于给定的每比特发射能量时，MIMO能支持更高的数据速率，所以它能以更快地速度发完，然后关闭电路以节省能量。

许多能量约束的网络由一些无法支持多天线的非常小的节点组成。此时，邻近节点相互协作可以形成一个多天线发射机或多天线的接收机，如图16-5所示。只要合作节点之间的距离较小，它们之间的信息交互所耗费能量就会小于所构成的MIMO系统节约的能量。文献[110]定量分析了合作MIMO的能量，结果指出，当发射和接收簇之间的距离是合作节点间距离的10到20倍时，这样的合作可以节省能量。当合作节点间的距离和这些簇的传输距离之间差不多时，本地信息交换花费的能量将超过了合作所节约的能量。合作MIMO是合作分集的一种。16.6.3节中讨论了其他一些技术，这些技术所能节省的能量与合作MIMO相当或更多。

560

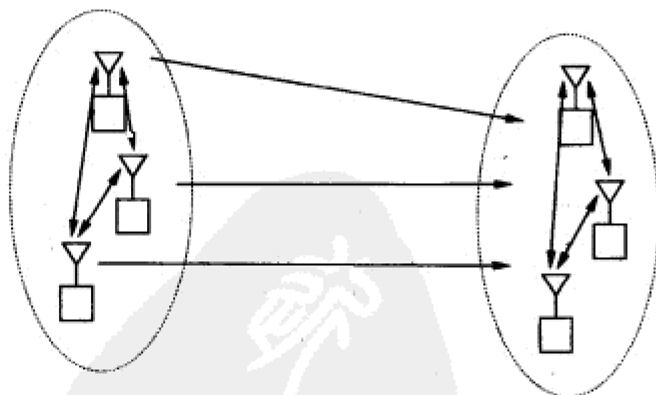


图16-5 合作MIMO

16.6.3 接入、路由和休眠

可以通过这样两种办法来提高随机接入的能量效率：(i) 尽量减少碰撞及由此产生的重传；(ii) 只用成功接收所需要的最小功率发射。减少碰撞的一种方法是当碰撞次

数增多时采用更强的差错保护[111]。在随机接入协议中通过探测来自适应地最小化功率也能显著提高能量效率[111, 38]。另一种方法是把分布式接入表示为博弈论问题，以能量和时延作为博弈中的代价[112]。文献[113]中评估了几种高能量效率的接入方法。不过没有哪一种协议能明显胜过其他，因为每种协议的性能都与信道特性密切相关。高能量效率的接入也可以纳入时延和公平性约束，文献[114]研究了这一点。这些技术中许多是以某种方式的TDMA来避免碰撞的，不过在分布式控制下，建立信道化接入的时延会比较大。

当用户有长串的分组包或者持续流数据时，随机接入的效率很差，因为大部分传输会发生碰撞。此时应当通过传输调度以系统化的方式把信道分配给用户。能量约束给调度优化带来了新的难题。根据文献[100]，使每比特的发送能量最小化的方法是将其在所有可用的带宽和时间维上发送。但有多个用户接入时，系统的时间和带宽资源必须分给这些用户。最近的研究探讨了多用户共享信道时，通过优化调度算法来最小化发射功率的问题[115]，它以截止期限或时延为约束条件，优化调度算法，以使每个用户所需要的发送能量最小化。优化能量的基础是根据时延约束合理地改变分组传输时间，改变传输时间就会改变能量消耗。在相同的截止期限约束下，这个方案比确定性调度有更高的能量效率。

能量约束的网络也要求路由协议能根据能量消耗优化路由。如果各节点消耗能量的速度不相同，有些节点的能量就会提前耗尽，致使网络出现了断裂。把每跳的代价从拥塞和时延改成能耗，就可以用16.3.3节所述的标准优化方法来得到端到端能耗最少的路由[116]。计算路由时也可以让代价和每个节点的电池相关联，例如尽量节约所有节点中电池寿命最小的那个电池的电量[116~117]。文献[116]通过仿真评估了能量约束路由优化中不同的代价函数，结果大致相同。也可以把代价函数加以扩展，使其除包含能量外，还包含传统的时延度量[118]。这一方法用不同的权重来控制时延和能耗对整体代价的贡献，从而使路由优化能在两者之间折中。注意，路由表的计算和发布也能造成显著的消耗，为此可按地理信息在目的地的大体方向上规划路由，这样做需要少量的预先计算[119]。 561

对于能量有限的节点，待机模式的消耗也是非常显著的，此时节点只是被动参与网络，有极少的数据交换以维持网络状态。几十年前寻呼业对这一问题的解决方案是让寻呼机周期性休眠。其基本想法是让寻呼机只在某个特定的短时间内监听传输。有中央控制器时，这个方法很容易实现。但在分布式控制的架构中怎么实现还不清楚。是否转入休眠还需要考虑到网络的连接性，这样的决定必须是本地非自主的。可在邻居发现的基础上结合某种排序判决做出是否休眠的决定。给定区域内，休眠机会应该在各节点中循环，以防止多个节点同时决定休眠而引起网络连接性问题。

16.6.4 能量约束下的跨层设计

能量约束网络的特性使跨层设计成为首选项。如果节点的电池不可充电，那么每个节点在其生存期内只能发送有限数目的比特，之后就不能再执行预定任务（如传感），也不能参加路由等网络活动。因此，必须在协议栈各层之间合理的使用能量，以延长网

络的生存期并满足业务要求。

协议栈的每一层一般都会引出能耗、时延和吞吐量之间的折中问题[120]。但对于某个给定的层，折中曲线上的最佳工作点应该考虑到高层设计。例如，某个节点也许会通过慢速发射来节约其能量，但这会使其他节点的接入问题变得复杂，并会增加端到端的时延。某个路由协议也许会考虑用中心节点来取得高能量效率的路由，但这会增加该路由上的拥塞和延迟，同时会很快耗尽这个节点的电池能量，致使其移出网络。能量、时延、吞吐量以及节点/网络寿命之间的折中最终应当针对业务要求进行优化。紧急援救行动需要快速获得现场信息，但这种支持本地信息交换的网络一般只需要持续几小时或几天。与此相反，浇注到桥梁中用来测量应力和张力的传感器网络需要能持续几十年，虽然信息收集只需要每天或每个星期进行一次。

16.6.5 单位能量的容量

发射能量受限时，不可能以无限小的错误率发送无限多个的比特。考虑单个比特的传输就能直观地理解这一点。对于信号空间中的两个点（代表一个比特的两个值），欲使其检测错误率充分小，只能让这两个点的空间分隔充分大，这要求充分大的能量。在硬能量约束下，任意小的差错率是不可能的，因此对于能量受限的节点，可靠通信需要不同的概念。

562

文献[121]（也见于文献[100]）提出的能量约束条件下可靠通信的容量定义为：差错概率随能量的增长而趋于零时，单位能量所能传输的最大比特数。这种单位能量容量的新概念要求在能量和分组长度趋于无限大时，差错率趋于零。因此按单位能量容量传输的有限能量系统不能使差错率无限小。为了理解这个概念的实质，我们来看AWGN信道下按归一化香农容量$C_B = C/B$bit/s/Hz传输所要求的每比特的最小能量[100]。当传输速率接近香农容量C时，每比特的接收能量为$E_b = P/R = P/C$。代入AWGN信道的香农容量公式可得

$$C = B\log_2\left(1 + \frac{P}{N_0 B}\right) = B\log_2\left(1 + \frac{E_b C}{N_0 B}\right) \tag{16-7}$$

对上式求逆可得到传输速率接近归一化容量$C_B = C/B$时要求的比特能量为

$$\frac{E_b}{N_0}(C_B) = \frac{2^{C_B} - 1}{C_B} \tag{16-8}$$

随着信道带宽B的增加，C_B将趋于零，于是宽带极限下最小比特能量为

$$\left(\frac{E_b}{N_0}\right)_{\min} = \lim_{C_B \to 0} \frac{2^{C_B} - 1}{C_B} = \ln 2 = -1.59\ \mathrm{dB} \tag{16-9}$$

文献[100]指出，一种类似于脉位调制的启闭式信号能在AWGN信道中达到这个最小的比特能量。此外，文献[122~123]的结果表明，可靠通信的最小比特能量对于平衰落信道也是式（16-9），即便接收机未知衰落。这些结果说明：(i) 无穷带宽时，最小比特能量不受衰落或接收机信息的影响，(ii) 对于最小能量通信，启闭式信号是近似最优的。

许多能量受限的无线系统带宽虽然很大，但还是有限的，无穷带宽下得到的结果有可能误导系统设计[123~124]。特别当比特能量接近于式（16-9）工作时，所需要的带宽对衰落分布和接收端对衰落信息的已知程度非常敏感。如果接收机已知衰落，则相干QPSK是最优的，启闭式信号显然是次优的。而在渐进条件下，接收机未知衰落时的启闭式信号也是最优的。

563

文献[121, 125~126]把单用户信道的单位能量容量扩展到了广播信道和多址信道。结果表明，在宽带极限下，TDMA对于这两种信道都是最优的，它能达到可靠通信所要求的最小比特能量。但在带宽很宽但有限的条件下，文献[126]指出，叠加方式（如CDMA结合多用户检测）能以和TDMA相同的最小比特能量达到可靠通信，但所需带宽更小。

习题

16.1　假设一个信号必须要传送1km远，路径损耗服从简化模型$P_r = P_t d^{-\gamma}$。

（a）若要求接收功率为10mW，分别计算出$\gamma = 2$和$\gamma = 4$时所需要的发射功率P_t。

（b）现在假设发射机和接收机之间的正中位置有一个中继节点，用它实现多跳路由。假设发射机和中继节点发射的功率都是P_t。当$\gamma = 2$和$\gamma = 4$时，为了使中继节点收到的来自发射机的功率为10mW，目的节点收到的自中继节点的功率也为10mW，问P_t应该是多大？并求出网络使用的总功率，即发射机和中继节点的功率总和。

（c）假设发射机和接收机之间均匀布置了N个中继节点，若要求每个中继节点以及接收机的接收功率都为10mW，请写出网络总功率的一般表达式。

16.2　某无线自组织网络有3个用户。用户1和用户2要求的SINR是7dB，用户3要求的SINR是10dB。假设所有接收机有相同的噪声功率$n_i = 1$，并且没有处理增益以减小干扰（$\rho = 1$）。假设增益矩阵为（行列序号表示用户编号）

$$G = \begin{bmatrix} 1 & 0.06 & 0.04 \\ 0.09 & 0.9 & 0.126 \\ 0.064 & 0.024 & 0.8 \end{bmatrix}$$

（a）证明方程$(\boldsymbol{I}-\boldsymbol{F})\boldsymbol{P} \geqslant \boldsymbol{u}$等价于各用户的SINR约束条件。

（b）证明此系统存在一个能满足所有SINR要求的功率向量。

（c）计算出能以最小的功率满足所有SINR要求的最优功率向量$\boldsymbol{P}^*$。

16.3　条件同上题。假设每个用户的初始功率为50，即初始功率向量为$\boldsymbol{P}(0)=(P_1(0), P_2(0), P_3(0)) = (50, 50, 50)$。每个用户按迭代公式（16-4）进行功控。请在$k = 1, \cdots, N$范围内画出每个用户（$i = 1, 2, 3$）的功率$P_i(k)$，其中$N$取足够大，以使功率向量$\boldsymbol{P}(N)$接近于最优值$\boldsymbol{P}^*$。同时画出当$k = 1, \cdots, N$时每个用户的SINR。

16.4　设网络节点沿一条直线分布在的无限网格上，节点间隔为10m。采用简化路径损耗模型$P_r = P_t d^{-\gamma}$，假设P_r至少需要10mW才能建立两个节点间的通信。

（a）当$\gamma = 2$时，计算使每个节点有$N = 2$个相邻节点的P_{max}。若节点的峰值功率小于P_{max}，将会有什么情况发生？

564

(b) $\gamma=2$、$N=4$时，计算P_{max}。

(c) $\gamma=4$、$N=4$时，计算P_{max}。

16.5 设有N个节点均匀分布在一个100m²的地理区域内，每个节点的发射功率足以使和距离R米内的其他节点通信。计算半径$R(1\leq R\leq 20)$内的平均节点数$E[N]$要求网络畅通。$E[N]$的平均值应通过范围内的100个随机采点的值计算。

16.6 在一个多路径路由的网络中，每个包从源节点出发，通过N个独立的路由发送到目的节点。假设每个端到端路由上的时延d是均值为D的指数分布：$p(d)=e^{-d/D}/D,\ d>0$。写出发送分组的N个副本到达目的节点的时延都超过D的概率与N的关系，并计算出N=1和N=5时的概率值。确定出N对网络吞吐量的影响。定量说明吞吐量和时延的折中关系。

16.7 证明式（16-6）对f_{ij}和C_{ij}都是上凸函数。

16.8 假设一条链路的容量是C = 10Mbit/s。请按数据流范围$0\leq f_{ij}\leq$ 10Mbit/s画出式（16-5）和式（16-6）给出的时延。

16.9 本题讨论两个时延度量式（16-5）和式（16-6）的区别。令λ为式（16-5）与式（16-6）的比值。

(a) 写出λ关于f_{ij}/C_{ij}的函数。

(b) 根据数据流f_{ij}必然小于C_{ij}这一点，给出λ的取值范围。

(c) 计算能使$\lambda>10$的f_{ij}/C_{ij}值。此时式（16-5）的时延度量比式（15-6）高一个数量级。

(d) 假设网络设计中分别用度量式（16-5）和度量式（16-6）计算路由代价，哪个度量下链路更为拥塞？为什么？

16.10 本题说明网络层和应用层之间的跨层设计所能带来的性能增益。对于视频类业务的传输，应用层总愿意用更高速率的编码方案以提高质量。现在假设业务是实时的。对于网络层决定的容量分配，如果传输速率高，链路上就会出现拥塞，而拥塞会造成时延，于是许多分组将不能按时到达译码器，导致通信质量变差。这就有了折中的问题。能反映这些效果的一个简单失真模型是

$$\mathrm{Dist}(R)=D_0+\frac{\theta}{R-R_0}+\kappa e^{-(C-R)T/L} \tag{16-10}$$

前两项对应应用层信源编码引起的失真，最后一项对应于分组时延引起的失真。令D_0= 0.38、R_0= 18.3kbit/s、θ = 2537，比例因子κ = 1，有效分组长度L = 3040bit，播放的截止期限T = 350ms。容量C和传输速率R在下面给出。

(a) 如果链路容量C分别以概率0.5、0.25和0.25取值于45kbit/s、24kbit/s和60kbit/s，计算能使平均失真Dist(R)最小的最优速率R。假设应用层和网络层之间完全合作，即应用层总知道当前网络层设置的容量C。

565

(b) 现在假定没有跨层优化，应用层总以固定的速率R = 22kbit/s编码。对于（a）中给出的容量分布，计算平均失真Dist(R)。

(c) 比较（a）和（b），算出无跨层优化时失真度增加的百分数。

16.11 式（16-8）给出的比值E_b/N_0是$C_B=C/B$的函数。证明：B固定时E_b/N_0随C增加，C固定时E_b/N_0随B减小。并证明 $\lim_{C_B\to 0}\left(\frac{E_b}{N_0}\right)(C_B)=\ln 2$。

参考文献

[1] R. E. Kahn, S. A. Gronemeyer, J. Burchfiel, and R. C. Kunzelman, "Advances in packet radio technology," *Proc. IEEE*, pp. 1468–96, November 1978.

[2] L. Kleinrock and J. Silvester, "Optimum transmission radii for packet radio networks or why six is a magic number," *Proc. IEEE Natl. Telecomm. Conf.*, pp. 4.3.1–4.3.5, December 1978.

[3] A. Ephremides, J. E. Wieselthier, and D. J. Baker, "A design concept for reliable mobile radio networks with frequency hopping signaling," *Proc. IEEE*, pp. 56–73, January 1987.

[4] J. Jubin and J. D. Tornow, "The DARPA packet radio network protocols," *Proc. IEEE*, pp. 21–32, January 1987.

[5] M. B. Pursley, "The role of spread spectrum in packet radio networks," *Proc. IEEE*, pp. 116–34, January 1987.

[6] F. A. Tobagi, "Modeling and performance analysis of multihop packet radio networks," *Proc. IEEE*, pp. 135–55, January 1987.

[7] B. Leiner, R. Ruther, and A. Sastry, "Goals and challenges of the DARPA Glomo program (global mobile information systems)," *IEEE Pers. Commun. Mag.*, pp. 34–43, December 1996.

[8] R. Ramanathan and R. Rosales-Hain, "Topology control of multihop wireless networks using transmit power adjustment," *Proc. IEEE Infocom Conf.*, pp. 404–13, March 2000.

[9] M. Ritter, R. Friday, and M. Cunningham, "The architecture of Metricom's microcellular data network and details of its implementation as the 2nd and 3rd generation ricochet wide-area mobile data service," *Proc. IEEE Emerg. Technol. Sympos. Broadband Commun.*, pp. 143–52, September 2001.

[10] M. N. Huhns, "Networking embedded agents," *IEEE Internet Comput.*, pp. 91–3, January/February 1999.

[11] W. W. Gibbs "As we may live," *Scientific American*, pp. 36, 40, November 2000.

[12] K. J. Negus, A. P. Stephens, and J. Lansford, "HomeRF: Wireless networking for the connected home," *IEEE Pers. Commun. Mag.*, pp. 20–7, February 2000.

[13] A. Schmidt, "How to build smart appliances," *IEEE Pers. Commun. Mag.*, pp. 66–71, August 2001.

[14] J. Haartsen, "The Bluetooth radio system," *IEEE Pers. Commun. Mag.*, pp. 28–36, February 2000.

[15] J. Haartsen and S. Mattisson, "Bluetooth: A new low-power radio interface providing short-range connectivity," *Proc. IEEE*, pp. 1651–61, October 2000.

[16] I. Poole, "What exactly is ... ZigBee?" *IEEE Commun. Eng.*, pp. 44–5, August/September 2004.

[17] T. Mitchell, "Broad is the way (ultra-wideband technology)," *IEE Review*, pp. 35–9, January 2001.

[18] J. Rabaey, J. Ammer, J. L. da Silva, Jr., and D. Roundy, "PicoRadio supports ad hoc ultra-low power wireless networking," *IEEE Computer*, pp. 42–8, July 2000.

566

[19] J. Nilsson, B. Bernhardsson, and B. Wittenmark, "Stochastic analysis and control of real-time systems with random time delays," *Automatica*, pp. 57–64, January 1998.

[20] X. Liu, S. S. Mahal, A. Goldsmith, and J. K. Hedrick, "Effects of communication delay on string stability in vehicle platoons," *Proc. IEEE Internat. Conf. Intell. Transp. Syst.*, pp. 625–30, August 2001.

[21] S. R. Graham, G. Baliga, and P. R. Kumar, "Issues in the convergence of control with communication and computing: Proliferation, architecture, design, services, and middleware," *Proc. IEEE Conf. Decision Control*, pp. 1466–71, December 2004.

[22] S. Basagni, D. Turgut, and S. K. Das, "Mobility-adaptive protocols for managing large ad hoc networks," *Proc. IEEE Internat. Conf. Commun.*, pp. 1539–43, June 2001.

[23] N. Bambos, "Toward power-sensitive network architectures in wireless communications: Concepts, issues, and design aspects," *IEEE Pers. Commun. Mag.*, pp. 50–9, June 1998.

[24] K. Sohrabi, J. Gao, V. Ailawadhi, and G. Pottie, "Protocols for self-organization of a wireless sensor network," *IEEE Pers. Commun. Mag.*, pp. 16–27, October 2000.

[25] L. Subramanian and R. H. Katz, "An architecture for building self-configurable systems," *Proc. Mobile Ad Hoc Network Comput. Workshop*, pp. 63–73, August 2000.

[26] R. Jain, A. Puri, and R. Sengupta, "Geographical routing using partial information for wireless ad hoc networks," *IEEE Pers. Commun. Mag.*, pp. 48–57, February 2001.

[27] L. Zhou and Z. J. Haas, "Securing ad hoc networks," *IEEE Network*, pp. 24–30, November/December 1999.

[28] R. Karri and P. Mishra, "Modeling energy efficient secure wireless networks using network simulation," *Proc. IEEE Internat. Conf. Commun.*, pp. 61–5, May 2003.

[29] *IEEE Wireless Commun. Mag.*, Special Issue on Energy Aware Ad Hoc Wireless Networks (A. J. Goldsmith and S. B. Wicker, Eds.), August 2002.

[30] S. Cui, A. J. Goldsmith, and A. Bahai, "Energy-constrained modulation optimization," *IEEE Trans. Wireless Commun.*, September 2005.

[31] D. Bertsekas and R. Gallager, *Data Networks*, 2nd ed., Prentice-Hall, Englewood Cliffs, NJ, 1992.

[32] V. Kawadia and P. R. Kumar, "Principles and protocols for power control in wireless ad hoc networks," *IEEE J. Sel. Areas Commun.*, pp. 76–88, January 2005.

[33] J. Zander, "Performance of optimum transmitter power control in cellular radio systems," *IEEE Trans. Veh. Tech.*, pp. 57–62, February 1992.

[34] S. A. Grandhi, R. Vijayan, and D. J. Goodman, "Distributed power control in cellular radio systems," *IEEE Trans. Commun.*, pp. 226–8, February–April 1994.

[35] G. J. Foschini and Z. Miljanic, "A simple distributed autonomous power control algorithm and its convergence," *IEEE Trans. Veh. Tech.*, pp. 641–6, November 1993.

[36] T. Holliday, N. Bambos, A. J. Goldsmith, and P. Glynn, "Distributed power control for time varying wireless networks: Optimality and convergence," *Proc. Allerton Conf. Commun., Control, Comput.*, pp. 1024–33, October 2003.

[37] N. Bambos, G. Pottie, and S. Chen, "Channel access algorithms with active link protection for wireless communications networks with power control," *IEEE/ACM Trans. Network.*, pp. 583–97, October 2000.

[38] S. Toumpis and A. J. Goldsmith, "Performance, optimization, and cross-layer design of media access protocols for wireless ad hoc networks," *Proc. IEEE Internat. Conf. Commun.*, pp. 2234–40, May 2003.

[39] T. ElBatt and A. Ephremides, "Joint scheduling and power control for wireless ad hoc networks," *IEEE Trans. Wireless Commun.*, pp. 74–85, January 2004.

[40] S. Kandukuri and N. Bambos, "Power controlled multiple access (PCMA) in wireless communication networks," *Proc. IEEE Infocom Conf.*, pp. 386–95, March 2000.

[41] S. Kallel, "Analysis of memory and incremental redundancy ARQ schemes over a nonstationary channel," *IEEE Trans. Commun.*, pp. 1474–80, September 1992.

[42] V. Rodoplu and T. H. Meng, "Minimum energy mobile wireless networks," *IEEE J. Sel. Areas Commun.*, pp. 1333–44, August 1999. 567

[43] P. Gupta and P. R. Kumar, "Towards an information theory of large networks: An achievable rate region," *IEEE Trans. Inform. Theory*, pp. 1877–94, August 2003.

[44] F. Xue and P. R. Kumar, "The number of neighbors needed for connectivity of wireless networks," *Wireless Networks*, pp. 169–81, March 2004.

[45] S. Toumpis and A. Goldsmith, "Capacity regions for ad hoc networks," *IEEE Trans. Wireless Commun.*, pp. 736–48, July 2003.

[46] O. Dousse, P. Thiran, and M. Hasler, "Connectivity in ad-hoc and hybrid networks," *Proc. IEEE Infocom Conf.*, pp. 1079–88, June 2002.

[47] B. Krishnamachari, S. B. Wicker, and R. Bejar, "Phase transition phenomena in wireless ad hoc

networks," *Proc. IEEE Globecom Conf.*, pp. 2921–5, November 2001.

[48] M. Penrose, *Random Geometric Graphs*, Oxford University Press, 2004.

[49] C.-K. Toh, *Ad Hoc Mobile Wireless Networks: Protocols and Systems*, Prentice-Hall, Englewood Cliffs, NJ, 2002.

[50] S.-J. Lee and M. Gerla, "Dynamic load-aware routing in ad hoc networks," *Proc. IEEE Internat. Conf. Commun.*, pp. 3206–10, June 2001.

[51] L. L. Peterson and B. S. Davie, *Computer Networks – A Systems Approach*, 2nd ed., Morgan-Kaufman, San Mateo, CA, 2000.

[52] C. P. P. Bhagwat, "Highly dynamic destination-sequenced distance vector routing (DSDV) for mobile computers," *Proc. ACM SIGCOMM*, pp. 234–44, September 1994.

[53] C. E. Perkins and E. M. Royer, "Ad hoc on-demand distance vector routing," *Proc. IEEE Workshop Mobile Comput. Syst. Appl.*, pp. 90–100, February 1999.

[54] D. B. Johnson and D. A. Maltz, "Dynamic source routing in ad hoc wireless networks," in T. Imielinsky and H. Korth, Eds., *Mobile Computing*, Kluwer, Dordrecht, 1996.

[55] M. R. Pearlman, Z. J. Haas, and S. I. Mir, "Using routing zones to support route maintenance in ad hoc networks," *Proc. IEEE Wireless Commun. Network Conf.*, pp. 1280–4, September 2000.

[56] A. Tsirigos and Z. J. Haas, "Multipath routing in the presence of frequent topological changes," *IEEE Commun. Mag.*, pp. 132–8, November 2001.

[57] D. Ayyagari, A. Michail, and A. Ephremides, "A unified approach to scheduling, access control and routing for ad-hoc wireless networks," *Proc. IEEE Veh. Tech. Conf.*, pp. 380–4, May 2000.

[58] S. Chen and K. Nahrstedt, "Distributed quality-of-service routing in ad hoc networks," *IEEE J. Sel. Areas Commun.*, pp. 1488–1505, August 1999.

[59] C. R. Lin and J.-S. Liu, "QoS routing in ad hoc wireless networks," *IEEE J. Sel. Areas Commun.*, pp. 1426–38, August 1999.

[60] M. Mirhakkak, N. Schult, and D. Thomson, "Dynamic bandwidth management and adaptive applications for a variable bandwidth wireless environment," *IEEE J. Sel. Areas Commun.*, pp. 1984–97, October 2001.

[61] A. Nosratinia, T. E. Hunter, and A. Hedayat, "Cooperative communication in wireless networks," *IEEE Commun. Mag.*, pp. 74–80, October 2004.

[62] S. Cui and A. J. Goldsmith, "Energy efficient routing using cooperative MIMO techniques," *Proc. IEEE Internat. Conf. Acous., Speech, Signal Proc.*, pp. 805–8, March 2005.

[63] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity. Part I: System description," *IEEE Trans. Commun.*, pp. 1927–38, November 2003.

[64] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity. Part II: Implementation aspects and performance analysis," *IEEE Trans. Commun.*, pp. 1939–48, November 2003.

[65] M. Janani, A. Hedayat, T. E. Hunter, and A. Nosratinia, "Coded cooperation in wireless communications: Space-time transmission and iterative decoding," *IEEE Trans. Signal Proc.*, pp. 362–71, February 2004.

[66] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Trans. Inform. Theory*, pp. 3062–80, December 2004.

[67] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung, "Network information flow," *IEEE Trans. Inform. Theory*, pp. 1204–16, July 2000.

568

[68] R. Koetter and M. Medard, "An algebraic approach to network coding," *IEEE/ACM Trans. Network.*, pp. 782–95, October 2003.

[69] S.-Y. R. Li, R. W. Yeung, and N. Cai, "Linear network coding," *IEEE Trans. Inform. Theory*, pp. 371–81, February 2003.

[70] N. Cai and R. W. Yeung, "Network error correction," *Proc. IEEE Inform. Theory Workshop.*, p. 101, June 2003.

[71] R. L. Cruz and A. V. Santhanam, "Optimal routing, link scheduling and power control in multihop wireless networks," *Proc. IEEE Infocom Conf.*, pp. 702–11, April 2003.
[72] M. Johansson, L. Xiao, and S. P. Boyd, "Simultaneous routing and power allocation in CDMA wireless data networks," *Proc. IEEE Internat. Conf. Commun.*, pp. 51–5, May 2003.
[73] L. Xiao, M. Johansson, and S. P. Boyd, "Optimal routing, link scheduling and power control in multihop wireless networks," *IEEE Trans. Commun.*, pp. 1136–44, July 2004.
[74] Y. Wu, P. A. Chou, Q. Zhang, K. Jain, W. Zhu, and S.-Y. Kung, "Network planning in wireless ad hoc networks: A cross-layer approach," *IEEE J. Sel. Areas Commun.*, pp. 136–50, January 2005.
[75] B. Radunović and J. Y. Boudec, "Rate performance objectives of multihop wireless networks," *IEEE Trans. Mobile Comput.*, pp. 334–49, October–December 2004.
[76] D. Julian, M. Chiang, D. O'Neill, and S. Boyd, "QoS and fairness constrained convex optimization of resource allocation for wireless cellular and ad hoc networks," *Proc. IEEE Infocom Conf.*, pp. 477–86, June 2002.
[77] Y. Wang and Q.-F. Zhu, "Error control and concealment for video communication: A review," *Proc. IEEE*, pp. 974–97, May 1998.
[78] E. Steinbach, N. Farber, and B. Girod, "Adaptive playout for low latency video streaming," *Proc. IEEE Internat. Conf. Image Proc.*, pp. 962–5, October 2001.
[79] D. G. Sachs, R. Anand, and K. Ramchandran, "Wireless image transmission using multiple-description based concatenated codes," *Proc. IEEE Data Compress. Conf.*, p. 569, March 2000.
[80] S. D. Servetto, K. Ramchandran, V. A. Vaishampayan, and K. Nahrstedt, "Multiple description wavelet based image coding," *IEEE Trans. Image Proc.*, pp. 813–26, May 2000.
[81] M. Alasti, K. Sayrafian-Pour, A. Ephremides, and N. Farvardin, "Multiple description coding in networks with congestion problem," *IEEE Trans. Inform. Theory*, pp. 891–902, March 2001.
[82] T. Holliday and A. Goldsmith, "Joint source and channel coding for MIMO systems: Is it better to be robust or quick," *Proc. Joint Workshop Commun. Cod.*, p. 24, October 2004.
[83] V. Kawadia and P. R. Kumar, "A cautionary perspective on cross-layer design," *IEEE Wireless Commun. Mag.*, pp. 3–11, February 2005.
[84] W. Kumwilaisak, Y. T. Hou, Q. Zhang, W. Zhu, C.-C. J. Kuo, and Y.-Q. Zhang, "A cross-layer quality-of-service mapping architecture for video delivery in wireless networks," *IEEE J. Sel. Areas Commun.*, pp. 1685–98, December 2003.
[85] T. Yoo, E. Setton, X. Zhu, A. Goldsmith, and B. Girod, "Cross-layer design for video streaming over wireless ad hoc networks," *Proc. IEEE Internat. Workshop Multi. Signal Proc.*, pp. 99–102, September 2004.
[86] W. Yu, K. J. R. Liu, and Z. Safar, "Scalable cross-layer rate allocation for image transmission over heterogeneous wireless networks," *Proc. IEEE Internat. Conf Acous., Speech, Signal Proc.*, pp. 4.593–4.596, May 2004.
[87] T. Cover and J. Thomas, *Elements of Information Theory*, Wiley, New York, 1991.
[88] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Trans. Inform. Theory*, pp. 388–404, March 2000.
[89] M. Grossglauser and D. N. C. Tse, "Mobility increases the capacity of ad-hoc wireless networks," *IEEE/ACM Trans. Network.*, pp. 1877–94, August 2002.
[90] A. El Gamal, J. Mammen, B. Prabhakar, and D. Shah, "Throughput–delay trade-off in wireless networks," *Proc. IEEE Infocom Conf.*, pp. 464–75, March 2004.
[91] S. Toumpis, "Large wireless networks under fading, mobility, and delay constraints," *Proc. IEEE Infocom Conf.*, pp. 609–19, March 2004.

569

[92] M. Gastpar and M. Vetterli, "On the capacity of wireless networks: The relay case," *Proc. IEEE Infocom Conf.*, pp. 1577–86, June 2002.
[93] A. Host-Madsen, "A new achievable rate for cooperative diversity based on generalized writing

on dirty paper," *Proc. IEEE Internat. Sympos. Inform. Theory*, p. 317, June 2003.

[94] A. Host-Madsen, "On the achievable rate for receiver cooperation in ad-hoc networks," *Proc. IEEE Internat. Sympos. Inform. Theory*, p. 272, June 2004.

[95] M. A. Khojastepour, A. Sabharwal, and B. Aazhang, "Improved achievable rates for user cooperation and relay channels," *Proc. IEEE Internat. Sympos. Inform. Theory*, p. 4, June 2004.

[96] N. Jindal, U. Mitra, and A. Goldsmith, "Capacity of ad-hoc networks with node cooperation," *Proc. IEEE Internat. Sympos. Inform. Theory*, p. 271, June 2004.

[97] C. T. K. Ng and A. J. Goldsmith, "Transmitter cooperation in ad-hoc wireless networks: Does dirty-paper coding beat relaying?" *IEEE Inform. Theory Workshop*, pp. 277–82, October 2004.

[98] A. Chandrakasan, R. Amirtharajah, S. Cho, J. Goodman, G. Konduri, J. Kulik, W. Rabiner, and A. Y. Wang, "Design considerations for distributed microsensor systems," *Proc. IEEE Custom Integrated Circuits Conf.*, pp. 279–86, May 1999.

[99] J. Rabaey, J. Ammer, J. L. da Silva, Jr., and D. Patel, "PicoRadio: Ad-hoc wireless networking of ubiquitous low-energy sensor/monitor nodes," *IEEE Comput. Soc. Workshop on VLSI*, pp. 9–12, April 2000.

[100] S. Verdú, "On channel capacity per unit cost," *IEEE Trans. Inform. Theory*, pp. 1019–30, September 1990.

[101] P. Agrawal, "Energy efficient protocols for wireless systems," *Proc. Internat. Sympos. Pers., Indoor, Mobile Radio Commun.*, pp. 564–9, September 1998.

[102] W. R. Heinzelman, A. Sinha, and A. P. Chandrakasan, "Energy-scalable algorithms and protocols for wireless microsensor networks," *Proc. IEEE Internat. Conf. Acous., Speech, Signal Proc.*, pp. 3722–5, June 2000.

[103] J. M. Kahn, R. H. Katz, and K. S. Pister, "Emerging challenges: Mobile networking for Smart Dust," *J. Commun. Networks*, pp. 188–96, August 2000.

[104] A. Chandrakasan and R. W. Brodersen, *Low Power Digital CMOS Design*, Kluwer, Norwell, MA, 1995.

[105] A. Ephremides, "Energy concerns in wireless networks," *IEEE Wireless Commun. Mag.*, pp. 48–59, August 2002.

[106] A. J. Goldsmith and S. B. Wicker, "Design challenges for energy-constrained ad hoc wireless networks," *IEEE Wireless Commun. Mag.*, pp. 8–27, August 2002.

[107] W. Stark, H. Wang. A. Worthen, S. Lafortune, and D. Teneketzis, "Low-energy wireless communication network design," *IEEE Wireless Commun. Mag.*, pp. 60–72, August 2002.

[108] E. Shih, P. Bahl, and M. Sinclair, "Wake on wireless: An event driven energy saving strategy for battery operated devices," *Proc. Internat. Conf. Mobile Comput. Network.*, pp. 160–71, September 2002.

[109] C. Schurgers and M. B. Srivastava, "Energy efficient wireless scheduling: Adaptive loading in time," *Proc. IEEE Wireless Commun. Network Conf.*, pp. 706–11, March 2002.

[110] S. Cui, A. J. Goldsmith, and A. Bahai, "Energy-efficiency of MIMO and cooperative MIMO in sensor networks," *IEEE J. Sel. Areas Commun.*, pp. 1089–98, August 2004.

[111] M. Zorzi and R. R. Rao, "Energy-constrained error control for wireless channels," *IEEE Pers. Commun. Mag.*, pp. 27–33, December 1997.

[112] A. B. MacKenzie and S. B. Wicker, "Selfish users in ALOHA: A game-theoretic approach," *Proc. IEEE Veh. Tech. Conf.*, pp. 1354–7, October 2001.

[113] A. Chockalingam and M. Zorzi, "Energy efficiency of media access protocols for mobile data networks," *IEEE Trans. Commun.*, pp. 1418–21, November 1998.

[114] J.-H. Youn and B. Bose, "An energy conserving medium access control protocol for multihop packet radio networks," *IEEE Internat. Conf. Comput. Commun. Network.*, pp. 470–5, October 2001.

570

[115] E. Uysal-Biyikoglu, B. Prabhakar, and A. El Gamal, "Energy-efficient packet transmission over

a wireless link," *IEEE/ACM Trans. Network.*, pp. 487–99, August 2002.

[116] C.-K. Toh, "Maximum battery life routing to support ubiquitous mobile computing in wireless ad hoc networks," *IEEE Commun. Mag.*, pp. 138–47, June 2001.

[117] J. H. Chang and L. Tassiulas, "Energy conserving routing in wireless ad-hoc networks," *Proc. IEEE Infocom Conf.*, pp. 609–19, August 2000.

[118] M. A. Youssef, M. F. Younis, and K. A. Arisha, "A constrained shortest-path energy-aware routing algorithm for wireless sensor networks," *Proc. IEEE Wireless Commun. Network Conf.*, pp. 794–9, March 2002.

[119] M. Zorzi and R. R. Rao, "Geographic random forwarding (GeRaF) for ad hoc and sensor networks: Energy and latency performance," *IEEE Trans. Mobile Comput.*, pp. 349–65, October–December 2003.

[120] S. Cui, R. Madan, A. J. Goldsmith, and S. Lall, "Joint routing, MAC, and link layer optimization in sensor networks with energy constraints," *Proc. IEEE Internat. Conf. Commun.*, May 2005.

[121] R. G. Gallager, "Energy limited channels: Coding, multiaccess and spread spectrum," *Proc. Inform. Syst. Sci. Conf.*, p. 372, March 1988. [See also Technical Report LIDS-P-1714, MIT, November 1987.]

[122] R. S. Kennedy, *Fading Dispersive Communication Channels*, Wiley, New York, 1969.

[123] S. Verdú, "Spectral efficiency in the wideband regime," *IEEE Trans. Inform. Theory*, pp. 1319–43, June 2002.

[124] S. Verdú, "Recent results on the capacity of wideband channels in the low-power regime," *IEEE Wireless Commun. Mag.*, pp. 40–5, August 2002.

[125] A. Lapidoth, I. E. Telatar, and R. Urbanke, "On wide-band broadcast channels," *IEEE Trans. Inform. Theory*, pp. 3250–8, December 2003.

[126] G. Caire, D. Tuninetti, and S. Verdú, "Suboptimality of TDMA in the low-power regime," *IEEE Trans. Inform. Theory*, pp. 608–20, April 2004.

571

附录A　带通信号和信道的表示

通信系统中的许多信号都是实带通信号，其频率响应以载波频率f_c为中心,覆盖一个窄带宽$2B$, $2B=f_c$的频带，如图A-1所示。由于是实带通信号，故其频率响应具有共轭对称性，即对于带通信号$s(t)$有$|S(f)|=|S(-f)|$，$\angle S(f)=-\angle S(-f)$。但带通信号不一定在其信号带宽内关于载频$f_c$共轭对称，也就是说，可能有某些$0<f\leqslant B$，使得$|S(f_c+f)|\neq|S(f_c-f)|$或者$\angle S(f_c+f)\neq-\angle S(f_c-f)$。图中示出了$|S(f)|$关于$f_c$的这种不对称性。带通信号或者是基带信号调制载波的结果，或者是确定或随机信号通过带通滤波器的结果。带通信号的带宽$2B$是f_c周围的一个大致的频率范围，此范围内信号的幅度不可忽略。通信系统中常常把带通信号作为发送信号和接收信号的模型。这些信号是实信号，因为发送电路只能产生实正弦信号（而不是复指数信号），信道的作用也只是改变实信号不同频率处的幅度和相位。

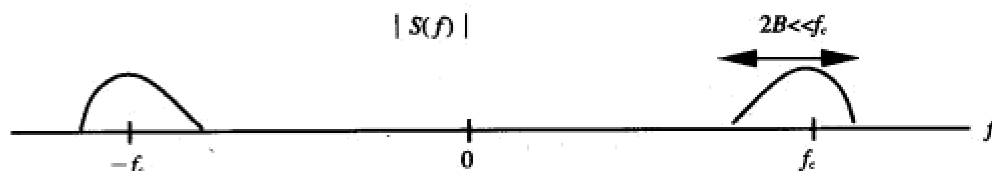


图　A-1

我们先把载频为f_c的带通信号$s(t)$表示为下面的形式：

$$s(t)=s_I(t)\cos(2\pi f_c t)-s_Q(t)\sin(2\pi f_c t) \tag{A-1}$$

其中，$s_I(t)$和$s_Q(t)$是带宽$B<<f_c$的实低通（基带）信号。这是带通信号或带通噪声的常用表示。MPSK、MQAM等调制就常常写成这样的形式。称$s_I(t)$为$s(t)$的**同相分量**(in-phase component)，$s_Q(t)$为$s(t)$的**正交分量**（quadrature component）。定义复信号$u(t)=s_I(t)+\mathrm{j}s_Q(t)$，则$s_I(t)=\mathrm{Re}\{u(t)\}$、$s_Q(t)=\mathrm{Im}\{u(t)\}$，于是$u(t)$是一个带宽为$B$的复基带信号。由此可得

$$s(t)=\mathrm{Re}\{u(t)\}\cos(2\pi f_c t)-\mathrm{Im}\{u(t)\}\sin(2\pi f_c t)=\mathrm{Re}\{u(t)\mathrm{e}^{\mathrm{j}2\pi f_c t}\} \tag{A-2}$$

上式右边的表达式称为带通信号$s(t)$的**复基带表示**（complex lowpass representation），基带信号$u(t)$称为$s(t)$的**等效基带信号**（equivalent lowpass signal）或者**复包络**（complex envelope）。注意，仅当$u(t)$是实信号，即$s_Q(t)=0$时，$U(f)$才是关于$f=0$共轭对称的。

根据傅里叶变换的性质可以证明：

$$S(f)=0.5[U(f-f_c)+U^*(-f-f_c)] \tag{A-3}$$

因为$s(t)$是实信号，所以$S(f)$关于$f=0$共轭对称。但基带信号$U(f)$和$U^*(f)$不一定关于$f=0$共轭对称，这一点将造成$S(f)$在带宽$2B$的范围内关于载频f_c不对称，见图A-1。实际上，仅当$u(t)=s_I(t)$，即$u(t)$中没有正交分量时，$S(f)$才在其带宽范围内关于载频共轭对称。稍后将看到，这种不对称会影响带通信道对带通信号的响应。

等效基带信号的另外一种表示为

PDG

$$u(t)=a(t)e^{j\phi(t)} \tag{A-4}$$

包络为

$$a(t)=\sqrt{s_I^2(t)+s_Q^2(t)} \tag{A-5}$$

相位为

$$\phi(t)=\tan^{-1}\left(\frac{s_Q(t)}{s_I(t)}\right) \tag{A-6}$$

根据这一表示，我们有

$$s(t)=\mathrm{Re}\{a(t)e^{j\phi(t)}e^{j2\pi f_c t}\}=a(t)\cos(2\pi f_c t+\phi(t)) \tag{A-7}$$

现在考虑实信道冲激响应$h(t)$，其傅里叶变换为$H(f)$。$h(t)$是实的，故$H^*(-f)=H(f)$。在通信系统中，我们主要关心$|f-f_c|<B$范围内的信道频率响应，因为只有$H(f)$的这些成分才会影响所关心的接收信号。**带通信道**（bandpass channel）类似于带通信号：实冲激响应$h(t)$的频率响应$H(f)$处在中心频率为f_c的$2B$带宽内，$2B<<f_c$。因为对应$H(f)$的冲激响应是一个带通信号，可以将其写成等效基带表示的形式：

$$h(t)=2\,\mathrm{Re}\{h_l(t)e^{j2\pi f_c t}\} \tag{A-8}$$

574

其中的系统2是为了避免在式（A-9）表示的$H(f)$中出现常系数因子。称$h_l(t)$为$H(f)$的**等效基带信道冲激响应**（equivalent lowpass channel impulse response）。根据式（A-2）和式（A-3），式（A-8）表明

$$H(f)=H_l(f-f_c)+H_l^*(-f-f_c) \tag{A-9}$$

于是$H(f)$由两部分组成：$H_l(f)$右移f_c，$H_l^*(f)$左移f_c。注意如果$H(f)$在带宽$2B$内关于载频f_c共轭对称，则$h_l(t)$是实函数，其频率响应$H_l(f)$将关于零共轭对称。但在很多无线信道（如频率选择性衰落信道）中，$H(f)$并不是关于f_c共轭对称的，此时$h_l(t)$是复函数，其同相分量为$h_{l,I}(t)=\mathrm{Re}\{h_l(t)\}$，正交分量为$h_{l,Q}(t)=\mathrm{Im}\{h_l(t)\}$。注意如果$h_l(t)$是复函数，则$H_l(f)$不是关于零共轭对称的。

现在我们用等效基带信号和等效基带信道模型来研究带通信号通过带通信道的输出。令$s(t)$表示输入信号，其等效基带信号为$u(t)$。令$h(t)$表示带通信道的冲激响应，其等效基带信道冲激响应为$h_l(t)$。发送信号$s(t)$和信道冲激响应$h(t)$都是实的，因而信道输出$r(t)=s(t)*h(t)$也是实的，其频域响应为$R(f)=H(f)S(f)$。由于$S(f)$是带通信号，因此$R(f)$也是带通信号，因此可以表示成复基带的形式：

$$r(t)=\mathrm{Re}\{v(t)e^{j2\pi f_c t}\} \tag{A-10}$$

现在我们来考虑信道输入$s(t)$、信道冲激响应$h(t)$、信道输出$r(t)$这三者的等效基带信号的关系。信道输出的频率响应可表示为：

$$\begin{aligned}R(f)&=H(f)S(f)\\&=0.5[H_l(f-f_c)+H_l^*(-f-f_c)][U(f-f_c)+U^*(-f-f_c)]\end{aligned} \tag{A-11}$$

对于带通信号和带通信道来说，$u(t)$和$h_l(t)$的带宽B远小于载频f_c，于是有

$$H_l(f-f_c)U^*(-f-f_c)=0$$

及

$$H_l^*(-f-f_c)U(f-f_c)=0$$

因此

$$R(f)=0.5[H_l(f-f_c)U(f-f_c)+H_l^*(-f-f_c)U^*(-f-f_c)] \tag{A-12}$$

根据式（A-2）和式（A-3），式（A-10）表明

$$R(f)=0.5[V(f-f_c)+V^*(-f-f_c)] \tag{A-13}$$

令式（A-12）和式（A-13）中的正频率部分和负频率部分相等，我们得到

$$V(f-f_c)=H_l(f-f_c)U(f-f_c) \tag{A-14}$$

及

575

$$V^*(-f-f_c)=H_l^*(-f-f_c)U^*(-f-f_c) \tag{A-15}$$

或者等价于

$$V(f)=H_l(f)U(f) \tag{A-16}$$

由傅里叶反变换可得：

$$v(t)=u(t)*h_l(t) \tag{A-17}$$

这样，通过$u(t)$和$h_l(t)$的卷积就可以得到接收信号$r(t)$的等效基带信号$v(t)$。接收信号因此是

$$r(t)=\mathrm{Re}\{(u(t)*h_l(t))e^{j2\pi f_c t}\} \tag{A-18}$$

注意仅当$H_l(f)$和$U(f)$都关于$f=0$共轭对称时，$V(f)=H_l(f)U(f)$关于$f=0$共轭对称。也就是说，如果$u(t)$或$h_l(t)$之一是复信号，则等效基带接收信号一般是同相分量和正交分量都不为零的复信号。此外，如果$u(t)=s_I(t)$是实信号（无正交分量），信道冲激响应$h_l(t)=h_{l,I}(t)+jh_{l,Q}(t)$是复信号（如频率选择性衰落），则

$$v(t)=s_I(t)*(h_{l,I}(t)+jh_{l,Q}(t))=s_I(t)*h_{l,I}(t)+js_I(t)*h_{l,Q}(t) \tag{A-19}$$

为复信号，因此接收信号将同时有同相分量和正交分量。更一般的情况下，如果$u(t)=s_I(t)+js_Q(t)$，且$h_l(t)=h_{l,I}(t)+jh_{l,Q}(t)$，则

$$\begin{aligned} v(t)&=[s_I(t)+js_Q(t)]*[h_{l,I}(t)+jh_{l,Q}(t)] \\ &=[s_I(t)*h_{l,I}(t)-s_Q(t)*h_{l,Q}(t)]+j[s_I(t)*h_{l,Q}(t)+s_Q(t)*h_{l,I}(t)] \end{aligned} \tag{A-20}$$

因此，$v(t)$的同相分量同时与$u(t)$的同相分量和正交分量有关。$v(t)$的正交分量也如此。这会对信号检测带来影响，它使调制信号的同相分量和正交分量在解调时相互干扰。

等效基带表示的主要目的是为了用发送信号、信道冲激响应和接收信号的等效基带模型来分析带通通信系统。这一方法在分析中消除了载波项，使分析不再依赖载波频率f_c。

576

附录B 概率论、随机变量和随机过程

本附录简要介绍书中所用到的概率论、随机变量、随机过程方面的主要概念。有关这一宽深主题的详细处理，以及本附录所给出的性质的证明，请参考文献[1~8]。

B.1 概率论

概率论是随机事件的数学描述。随机事件由概率空间$(\Omega, \varepsilon, p(\cdot))$定义。概率空间由样本空间$\Omega$、随机事件的集合$\varepsilon$和概率测度$p(\cdot)$组成。其中$\Omega$是随机事件可能结果的集合。$\varepsilon$是集合的集合，任意随机事件$A\in\varepsilon$是$\Omega$的子集。对每一个集合$A\in\varepsilon$定义了概率测度$p(A)$。概率空间要求集合$\varepsilon$是一个$\sigma$域。直观地说，如果一个集合的$\varepsilon$集合$R$包含了所有它的元素的交集、并集和补集[1]，$\varepsilon$就是一个$\sigma$域。更准确地说，$\varepsilon$是一个$\sigma$域，如果：所有可能结果构成的集合$\Omega$是$\varepsilon$中的一个集合；若集合$A\in\varepsilon$，则$A^c\in\varepsilon$；对于任意集合$A_1, A_2, \cdots$，其中$A_i\in\varepsilon$，有$\bigcup_{i=1}^{\infty} A_i\in\varepsilon$。为了能定义随机事件的交、并的概率，$\varepsilon$必须是$\sigma$域。我们还要求概率空间中的概率测度满足下列三个基本性质：

(1) $p(\Omega)=1$。

(2) 对于任意事件$A\in\varepsilon$，有$0\leqslant p(A)\leqslant 1$。

(3) 如果A和B是互斥的（即其交集为零），则$p(A\cup B)=p(A)+p(B)$。

本节只考虑ε中的集合，因为概率测度只定义在这些集合上。

从概率测度$p(\cdot)$的基本性质可以推出一些重要特性，如$p(A^c)=1-p(A)$。再如，若集合$A_1, \cdots, A_n$两两不相交（$A_i\cap A_j=\emptyset$, $i\neq j$），则当$A_1\cup A_2\cup\cdots\cup A_n=\Omega$时，有$\sum_{i=1}^{n}p(A_i)=1$，称这样的集合$\{A_1, \cdots, A_n\}$为$\Omega$的一个划分（partition）。对于两个相交的集合A_i和A_j，有$p(A_i\cup A_j)=p(A_i)+p(A_j)-p(A_i\cap A_j)$，这一点导出了联合界（union bound），其表述为：对于任意集合$A_1, \cdots, A_n$，有

$$p(A_1\cup A_2\cup\cdots\cup A_n)\leqslant\sum_{i=1}^{n}p(A_i) \tag{B-1}$$

一个随机事件的发生能够影响另外一个随机事件发生的概率，这是因为观察到一个随机事件的观察结果后，我们能够判定出ε中有哪些子集也包括这个观察结果。为了反映这一点，定义事件B在事件A发生的条件下的概率为$p(B|A)=p(A\cap B)/p(A)$，设$p(A)\neq 0$。这表明

$$p(A\cap B)=p(A\mid B)p(B)=p(B\mid A)p(A) \tag{B-2}$$

条件概率$p(B|A)=p(A\cap B)/p(A)$实际是用事件A的概率对事件B的概率进行了归一化，因

1. 我们用$A\cap B$表示A和B的交集，它是所有A和B的共同元素。A和B的并集记为$A\cup B$，是所有或出现在A中、或出现在B中的元素的集合。$A\subset\Omega$的补集记为A^c，是所有在Ω中，但不在A中的元素。

为我们知道A已经发生了。由（B-2）可得到贝叶斯准则（Bayes'rule）：

$$p(B \mid A)=\frac{p(A \mid B)p(B)}{p(A)} \tag{B-3}$$

事件的独立性与概率测度$p(\cdot)$有关，若$p(A\cap B)=p(A)p(B)$，则事件A与事件B独立。此时有$p(B|A)=p(B)$，$p(A|B)=p(A)$。

B.2 随机变量

随机变量是在概率空间$(\Omega, \varepsilon, p(\cdot))$上定义的。随机变量$X$是从样本空间$\Omega$到实数轴的子集的函数映射。如果$X$取实数轴上的离散值，称为离散（discrete）随机变量。如果X取实数轴上的连续值，称为连续（continuous）随机变量。随机变量X的累积分布函数（cumulative distribution function，cdf）定义为$P_X(x)\triangleq p(X\leqslant x)$，$x\in\mathbb{R}$。累积分布函数可以从概率空间导出：$p(X\leqslant x)=p(X^{-1}(-\infty, x))$，$X^{-1}(\cdot)$是从实数轴到$\Omega$的子集的逆映射，即$X^{-1}(-\infty, x)=\{\omega\in\Omega: X(\omega)\leqslant x\}$。累积分布函数的性质基于概率测度的性质，首先它满足$0\leqslant P_X(x)=p(X^{-1}(-\infty, x))\leqslant 1$。其次，累积分布函数是不减函数：如果$x_1\leqslant x_2$，则$P_X(x_1)\leqslant P_X(x_2)$，这是由于$P_X(x_2)=p(X^{-1}(-\infty, x_2))=p(X^{-1}(-\infty, x_1))+p(X^{-1}(x_1, x_2))\geqslant p(X^{-1}(-\infty, x_1))=P_X(x_1)$。

随机变量X的概率密度函数（probability density function，pdf）定义为累积分布函数的导数：$p_X(x)\triangleq\frac{\mathrm{d}}{\mathrm{d}x}P_X(x)$。对连续随机变量，$p_X(x)$是整个实数轴上的函数。对离散随机变量，$p_X(x)$是一组冲激函数，冲激位置在$X$的可能取值处。概率密度函数也称为$X$的概率分布（probability distribution）或分布（distribution），它决定了X处于某一范围时的概率：

$$p(x_1<X\leqslant x_2)=p(X\leqslant x_2)-p(X\leqslant x_1)=P_X(x_2)-P_X(x_1)=\int_{x_1}^{x_2}p_X(x)\,\mathrm{d}x \tag{B-4}$$

由于$P_X(\infty)=1$、$P_X(-\infty)=0$，所以概率密度函数的积分是1：

$$\int_{-\infty}^{\infty}p_X(x)\,\mathrm{d}x=1 \tag{B-5}$$

在不致混淆的情况下，可以省略累积分布函数和概率密度函数中的下标X，写成$P(x)$和$p(x)$。

578

随机变量X的均值（mean）或期望值（expected value）是其概率平均，定义为：

$$\mu_X=E[X]\triangleq\int_{-\infty}^{\infty}xp_X(x)\,\mathrm{d}x \tag{B-6}$$

期望算子$E[\cdot]$是线性的，也可用于随机变量的函数。X的函数的均值为

$$E[g(X)]=\int_{-\infty}^{\infty}g(x)p_X(x)\,\mathrm{d}x \tag{B-7}$$

一个有特别意义的函数是X的n阶矩（moment）：

$$E[X^n]=\int_{-\infty}^{\infty}x^n p_X(x)\,\mathrm{d}x \tag{B-8}$$

X的方差是由其均值和二阶矩定义的：

$$\mathrm{Var}[X]=\sigma_X^2 \triangleq E[(X-\mu_X)^2]=E[X^2]-\mu_X^2 \tag{B-9}$$

方差反映X与其均值μ_X之差的平方的均值。X的标准差σ_X是方差的平方根。由期望算子的线性性质容易证明，对任意常数c，有$E[cX]=cE[X]$、$\mathrm{Var}[cX]=c^2\mathrm{Var}[X]$、$E[X+c]=E[X]+c$、$\mathrm{Var}[X+c]=\mathrm{Var}[X]$。因此，给一个随机变量乘以一个常数将使其均值乘以相同的常数，使其方差乘以该常数的平方。给一个随机变量加上一个常数将使均值加上相同的常数，而方差不变。

随机变量X的分布可以通过它的**特征函数**（characteristic function）来确定，特征函数定义为：

$$\phi_X(\nu)\triangleq E[\mathrm{e}^{\mathrm{j}\nu X}]=\int_{-\infty}^{\infty}p_X(x)\mathrm{e}^{\mathrm{j}\nu x}\,\mathrm{d}x \tag{B-10}$$

由（B-10）可见，X的特征函数$\phi_X(\nu)$是其概率密度函数$p_X(x)$的傅里叶反变换在$\nu/(2\pi)$处的值。因此通过$\phi_X(\nu)$可得到$p_X(x)$为：

$$p_X(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\phi_X(\nu)\mathrm{e}^{-\mathrm{j}\nu x}\,\mathrm{d}x \tag{B-11}$$

579

上式在求随机变量之和的分布时特别有用。可以由$\phi_X(\nu)$得到X的n阶矩：

$$E[X^n]=(-\mathrm{j})^n\left.\frac{\partial^n\phi_X(\nu)}{\partial\nu^n}\right|_{\nu=0}$$

X的**矩母函数**（moment generating function，MGF）定义为$\mathcal{M}_X(\nu)\triangleq E\left[\mathrm{e}^{\nu X}\right]$，它与特征函数类似，但在某些$\nu$值处会发散。如果矩母函数在零附近是有限的，则$X$的$n$阶矩为：

$$E[X^n]=\left.\frac{\partial^n\mathcal{M}_X(\nu)}{\partial\nu^n}\right|_{\nu=0}$$

令X是一个随机变量，$g(x)$是一个实函数。令$Y=g(X)$就定义了另外一个随机变量，且有$P_Y(y)=\int_{x:g(x)\leqslant y}p_X(x)\mathrm{d}x$。若$g$是一一映射的单调增函数，则$P_Y(y)=\int_{-\infty}^{g^{-1}(y)}p_X(x)\mathrm{d}x$。若$g$是一一映射的单调减函数，则$P_Y(y)=\int_{g^{-1}(y)}^{\infty}p_X(x)\mathrm{d}x$。

现在考虑联合的随机变量。为了能定义两个随机变量的联合分布，它们必须有相同的概率空间。令X和Y是定义在同一个概率空间$(\Omega,\ \varepsilon,\ p(\cdot))$上的两个随机变量。它们的联合累积分布函数定义为$P_{XY}(x,y)\triangleq p(X\leqslant x,\ Y\leqslant y)$。联合概率密度函数定义为累积分布函数的导数：

PDG

$$p_{XY}(x,y) \triangleq \frac{\partial^2 P_{XY}(x,y)}{\partial x \partial y} \tag{B-12}$$

于是，

$$P_{XY}(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} p_{XY}(v,w)\,\mathrm{d}v\,\mathrm{d}w \tag{B-13}$$

对于联合随机变量X和Y，对联合概率密度函数求关于Y的积分可得到X的分布：

$$p_X(x) = \int_{-\infty}^{\infty} p_{XY}(x,y)\,\mathrm{d}y \tag{B-14}$$

类似地，

$$p_Y(y) = \int_{-\infty}^{\infty} p_{XY}(x,y)\,\mathrm{d}x \tag{B-15}$$

这样得到的分布$p_X(x)$和$p_Y(y)$也称为联合分布$p_{XY}(x, y)$的边际（marginal）分布。注意联合概率密度函数的积分必然是1：

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p_{XY}(x,y)\,\mathrm{d}x\,\mathrm{d}y = 1 \tag{B-16}$$

两个随机变量的联合累积分布函数和联合概率密度函数的定义可以直接扩展到任意有限个随机变量。

580

和随机事件一样，观察一个随机变量的结果可能会影响另一个随机变量的概率。在随机变量X的实现给定为$X=x$的条件下，随机变量Y的条件分布定义为$p_Y(y|X = x)=p_{XY}(x, y)/p_X(x)$，这也表明$p_{XY}(x, y)=p_Y(y|X = x)p_X(x)$。两个随机变量$X$和$Y$之间的独立性是其联合分布的函数。具体而言，若$X$和$Y$的联合分布$p_{XY}(x, y)$可分解为各自分布之积，即若$p_{XY}(x, y)=p_X(x)p_Y(y)$，则$X$和$Y$是独立的随机变量。对于独立随机变量，容易证明$\boldsymbol{E}[f(X)g(Y)]=\boldsymbol{E}[f(X)]\boldsymbol{E}[g(Y)]$，其中$f(x)$和$g(x)$是任意函数。

设随机变量X和Y的联合概率密度函数是$p_{XY}(x, y)$，定义其ij阶联合矩为：

$$\boldsymbol{E}[X^iY^j] \triangleq \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^i y^j p_{XY}(x,y)\,\mathrm{d}x\,\mathrm{d}y \tag{B-17}$$

X和Y的*相关*（correlation）定义为$\boldsymbol{E}[XY]$，*协方差*（covariance）定义为$\mathrm{Cov}[XY] \triangleq \boldsymbol{E}[(X-\mu_X)(Y-\mu_Y)]=\boldsymbol{E}[XY]-\mu_X\mu_Y$。注意若$X$和$Y$当中有一个是零均值，则它们的协方差和相关值相等。$X$和$Y$的*相关系数*（correlation coefficient）由其协方差和标准差定义为$\rho \triangleq \mathrm{Cov}[XY]/(\sigma_X\sigma_Y)$。若$X$和$Y$的协方差为零，或等效地说是若其相关系数为零，则称它们不*相关*（uncorrelated）。注意，对于两个不相关的随机变量（即$\mathrm{Cov}[XY]=\boldsymbol{E}[XY]-\mu_X\mu_Y=0$），若其均值不为零，则相关函数也不为零($\boldsymbol{E}[XY]\neq 0$)。对于随机变量$X_1,\cdots, X_n$，其*协方差矩阵*（covariance matrix）Σ定义为一个$n\times n$的矩阵，第ij个元素是$\Sigma_{ij} = \mathrm{Cov}[X_iY_j]$。$\Sigma$对角线上的第$i$个元素是$X_i$的方差：$\Sigma_{ii} = \mathrm{Var}[X_i]$。

考虑两个独立随机变量X和Y。令$Z=X+Y$就定义一个概率空间$(\Omega, \varepsilon, p(\cdot))$上新随机变

量。用特征函数可以证明，Z的分布是X和Y的分布的卷积：$p_Z(z)=p_X(x)*p_Y(y)$，等效于$\phi_Z(v)=\phi_X(v)\phi_Y(v)$。根据这个分布可以导出$E[Z]=E[X]+E[Y]$、$\text{Var}[Z]=\text{Var}[X]+\text{Var}[Y]$。因此，独立随机变量之和的均值为均值的和，和的方差为方差的和。

研究通信系统的时候经常出现的一个分布是高斯分布。随机变量X的高斯分布由其均值μ_X和方差σ_X^2定义：

$$p_X(x)=\frac{1}{\sqrt{2\pi}\sigma_X}\mathrm{e}^{-[(x-\mu_X)^2/(2\sigma_X^2)]} \tag{B-18}$$

高斯分布也称正态分布，记为$N(\mu_X,\sigma_X^2)$。注意高斯分布的拖尾（即x远离μ_X时$p_X(x)$的值）是指数下降的。高斯分布的累积分布函数$P_X(x)=p(X\leqslant x)$无闭式解，可用高斯Q函数表示：

$$P_X(x)=p(X\leqslant x)=1-Q\left(\frac{x-\mu_X}{\sigma_X}\right) \tag{B-19}$$

其中的高斯Q函数定义为：

$$Q(x)\triangleq\int_x^\infty\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-y^2/2}\,\mathrm{d}y \tag{B-20}$$

它是均值为零、方差为1的高斯随机变量$X\sim N(0,1)$大于x的概率：$Q(x)=p(X\geqslant x)$。高斯Q函数与互补误差函数的关系为：$Q(x)=0.5\text{erfc}\left(x/\sqrt{2}\right)$。这些函数一般可用标准的计算机数学程序包计算。

581

令$\boldsymbol{X}=(X_1,\cdots,X_n)$表示联合高斯随机变量组成的向量，它们的联合分布为：

$$p_{X_1,\cdots,X_n}(x_1,\cdots,x_n)=\frac{1}{\sqrt{(2\pi)^n\det[\boldsymbol{\Sigma}]}}\exp[-0.5(X-\mu_X)^T\boldsymbol{\Sigma}^{-1}(X-\mu_X)] \tag{B-21}$$

其中$\mu_X=\boldsymbol{E}[X]^T=(\boldsymbol{E}[X_1],\cdots,\boldsymbol{E}[X_n])^T$是$\boldsymbol{X}$的均值，$\Sigma$是$\boldsymbol{X}$的$n\times n$协方差矩阵，即$\Sigma_{ij}=\text{Cov}[X_iX_j]$。根据（B-21）可以证明，对于联合高斯随机变量$X$和$Y$，若$\text{Cov}[XY]=0$则$p_{XY}(x,y)=p_X(x)p_Y(y)$。也就是说，高斯随机变量如果不相关就是独立的。

一个复随机变量Z是复高斯的，如果$Z=X+\mathrm{j}Y$，其中X和Y是实联合高斯随机变量。Z的分布即由X和Y的联合分布给定，即式（B-21）中矢量为(X,Y)。类似地，复随机矢量$\mathbf{Z}=(Z_1,\cdots,Z_N)=(X_1+\mathrm{j}Y_1,\cdots,X_N+\mathrm{j}Y_N)$是复高斯的，如果$X_1,\cdots,X_N,Y_1,\cdots,Y_N$是联合高斯的实随机变量，$\mathbf{Z}$的分布由$X_1,\cdots,X_N,Y_1,\cdots,Y_N$的联合分布给定，即式（B-21）中矢量为$(X_1,\cdots,X_N,Y_1,\cdots,Y_N)$。

通信系统模型中经常出现高斯分布的原因是*中心极限定理*（central limit theorem，CLT）。这个定理给出了大量独立同分布随机变量之和的极限分布。设X_i是独立同分布的联合随机变量，令$Y_n=\sum_{i=1}^n X_i$、$Z_n=(Y_n-\mu_{Y_n})/\sigma_{Y_n}$。中心极限定理指出，随着$n$趋向无穷，$Z_n$的分布收敛于均值为0、方差为1的高斯分布，即$\lim_{n\to\infty}p_{Z_n}(x)=N(0,1)$。因此，任何随机变量，若它等于大量独立同分布随机变量之和，则它的分布近似于高斯分布。例如无

线接收机中的噪声是由各种硬件产生的无用信号组成的，大量的独立同分布成分使高斯分布模型能准确地表示这种噪声。

通信系统中常见的另外两种分布是均匀分布和二项分布。均匀分布随机变量X的概率密度函数在$x\in[a, b]$时为$p_X(x) = 1/(b-a)$，x取其他值时为零。随机相位θ的模型一般是$[0, 2\pi]$内的均匀分布，记为$\theta\sim\mathcal{U}[0, 2\pi]$。二项分布常出现在编码分析中。令$X_i(i=1, \cdots, n)$是取值于0或1的离散随机变量，假设$X_i$为独立同分布，且$p(X_i = 1) = p$、$p(X_i = 0) = 1-p$。令$Y=\sum_{i=1}^{n}X_i$，则$Y$是取值为整数$k=0,1,2,\cdots,n$的离散随机变量。$Y$的分布就是二项分布，为：

$$p(Y=k)=\binom{n}{k}p^k(1-p)^{n-k} \tag{B-22}$$

其中

582

$$\binom{n}{k}\triangleq\frac{n!}{k!\,(n-k)!} \tag{B-23}$$

B.3 随机过程

随机过程$X(t)$定义在概率空间$(\Omega, \varepsilon, p(\cdot))$之上。它是从样本空间$\Omega$到实函数集$\{x_1(t), x_2(t),\cdots\}$的映射，$x_i(t)$是$X(t)$的可能实现。$X(t)$在时刻$t_0, t_1, \cdots, t_n$的采样值是定义在这个概率空间上的联合随机变量。因而这些采样的联合分布函数为$P_{X(t_0)X(t_1)\cdots X(t_n)}(x_0, \cdots, x_n) = p(X(t_0)\leqslant x_0, X(t_1)\leqslant x_1, \cdots, X(t_n)\leqslant x_n)$。随机过程$X(t)$的特性由任意抽样时刻$\{t_0, t_1, \cdots, t_n\}$的采样值的联合分布$P_{X(t_0)X(t_1)\cdots X(t_n)}(x_0, \cdots, x_n)$完全确定。

随机过程$X(t)$是*平稳*（stationary）的，若对于任意的T，任意的n，以及任意的采样时刻$\{t_0, t_1, \cdots, t_n\}$，有

$$\begin{aligned}&p\big(X(t_0)\leqslant x_0, X(t_1)\leqslant x_1,\cdots,X(t_n)\leqslant x_n\big)\\&\qquad=p\big(X(t_0+T)\leqslant x_0, X(t_1+T)\leqslant x_1,\cdots,X(t_n+T)\leqslant x_n\big)\end{aligned}$$

直观来说，如果时移并不影响随机过程的概率特性，它就是平稳的。随机过程的平稳性一般难以证明，因为它要求检验所有可能的时移下，所有可能采样的联合分布。一般通过产生随机过程的源的特性来推断是否具有平稳性。

随机过程的*均值*（mean）定义为$E[X(t)]$。由于平稳过程的均值独立于时移，所以它一定是常数：$E[X(t)]=E[X(t-t)]=E[X(0)]=\mu_X$。随机过程的*自相关函数*（autocorrelation）定义为$A_X(t, t+\tau)\triangleq E[X(t)X(t+\tau)]$，也称为$X(t)$的*二阶矩*（second moment）。由于平稳过程的自相关函数独立于时移，所以$A_X(t, t+\tau)=E[X(t-t)X(t+\tau-t)]=E[X(0)X(\tau)]\triangleq A_X(\tau)$。因此，平稳过程的自相关函数只与样值$X(t)$和$X(t+\tau)$的采样时间差$\tau$有关，与绝对时间$t$无关。自相关函数反映随机过程在不同时间的采样之间的相关性。

定义同一概率空间上的两个随机过程$X(t)$和$Y(t)$有如下的联合分布函数：

$$\begin{aligned}&P_{X(t_0),\cdots,X(t_n)Y(t'_0),\cdots,Y(t'_m)}(x_0,\cdots,x_n,y_0,\cdots,y_m)\\&\qquad=p\big(X(t_0)\leqslant x_0,\cdots,X(t_n)\leqslant x_n,Y(t'_0)\leqslant y_0,\cdots,Y(t'_m)\leqslant y_m\big)\end{aligned} \tag{B-24}$$

其中$\{t_0, t_1, \cdots, t_n\}$和$\{t'_0, t'_1, \cdots, t'_n\}$是任意的采样时刻。$X(t)$和$Y(t)$是*独立*（independent）的，若对这些采样时刻有：

$$p_{X(t_0),\cdots,X(t_n)Y(t'_0),\cdots,Y(t'_m)}(X(t_0) \leqslant x_0, \cdots, X(t_n) \leqslant x_n, Y(t'_0) \leqslant y_0, \cdots, Y(t'_m) \leqslant y_m)$$
$$= p_{X(t_0),\cdots,X(t_n)}(X(t_0) \leqslant x_0, \cdots, X(t_n) \leqslant x_n)$$
$$\cdot p_{Y(t'_0),\cdots,Y(t'_m)}(Y(t'_0) \leqslant y_0, \cdots, Y(t'_m) \leqslant y_m) \quad \text{(B-25)}$$

随机过程$X(t)$和$Y(t)$的互相关函数定义为

$$A_{XY}(t, t+\tau) \triangleq E[X(t)Y(t+\tau)]$$

若对任意的t和τ，有$E[X(t)Y(t+\tau)]=E[X(0)]E[Y(t+\tau)]$，则这两个过程是不相关的。若$X(t)$和$Y(t)$都是平稳过程，则互相关函数只与$\tau$有关：$A_{XY}(t, t+\tau)=E[X(t-t)Y(t+\tau-t)]=E[X(0)Y(\tau)] \triangleq A_{XY}(\tau)$。

583

许多分析只涉及一阶矩和二阶矩，而*广义平稳*（wide-sense stationary，WSS）便是一种只涉及一二阶矩的平稳概念，并且也容易验证。若一个随机过程的均值为常数、自相关只与采样的时间差有关，即若$E[X(t)]=\mu_X$、$A_X(t, t+\tau) = E[X(t)X(t+\tau)]=A_X(\tau)$，则此过程为广义平稳的。平稳过程是广义平稳的，但广义平稳过程不一定是平稳的。广义平稳过程的自相关函数是τ的偶函数，即$A_X(\tau)=E[X(t)X(t+\tau)]=E[X(t+\tau)X(t)]=A_X(-\tau)$，并且$A_X(\tau)$在$\tau=0$时最大，即$|A_X(\tau)| \leqslant A_X(0)=E[X^2(t)]$。和平稳过程一样，若$X(t)$和$Y(t)$均为广义平稳，则它们的互相关函数与时移无关，只与这两个过程的时间差有关：$A_{XY}(t, t+\tau)=E[X(0)Y(\tau)]=A_{XY}(\tau)$。

广义平稳过程的*功率谱密度*（power spectral density，PSD）定义为自相关函数关于τ的傅里叶变换：

$$S_X(f) = \int_{-\infty}^{\infty} A_X(\tau)e^{-j2\pi f\tau}\,d\tau \quad \text{(B-26)}$$

对功率谱密度做反变换可也得到自相关函数：

$$A_X(\tau) = \int_{-\infty}^{\infty} S_X(f)e^{j2\pi f\tau}\,df \quad \text{(B-27)}$$

叫功率谱密度是因为它的积分就是随机过程$X(t)$的平均功率。由式（B-27）可得

$$E[X^2(t)] = A_X(0) = \int_{-\infty}^{\infty} S_X(f)\,df \quad \text{(B-28)}$$

同样由式（B-26）可得$S_X(0)=\int_{-\infty}^{\infty} A_X(\tau)d\tau$。由于$A_X(\tau)$是实偶函数，所以由（B-26）可知$S_X(f)$也是偶函数，即$S_X(f)=S_X(-f)$。*白噪声*（white noise）定义为一个零均值的广义平稳随机过程，其功率谱密度对所有频率都是常数。即白噪声过程$X(t)$有$E[X(t)]=0$和$S_X(f)=N_0/2$，常数N_0称为白噪声的单边功率谱密度。通过傅里叶反变换可得白噪声的自相关函数为$A_X(\tau)=(N_0/2)\delta(\tau)$。从某种意义来说，白噪声是所有噪声中最随机的，它经过一个瞬时就变得不相关了。

经常要对随机过程进行滤波或者调制，广义平稳性会使相关的问题便于分析。功率谱密度为$S_X(f)$的广义平稳过程通过线性时不变系统$H(f)$的输出也是广义平稳的，其功率谱密度为$|H(f)|^2 S_X(f)$。若将此$X(t)$与载波$\cos(2\pi f_c t+\theta)$相乘，其中$\theta \sim \mathcal{U}[0, 2\pi]$，则相乘的结果$X(t)\cos(2\pi f_c t+\theta)$也是广义平稳的，其功率谱密度为$0.25[S_X(f-f_c)+S_X(f+f_c)]$。

584

将一个均值为零、功率谱密度为$N_0/2$的白高斯噪声过程通过一个中心频率为f_c、带宽$2B$的理想带通滤波器，其输出是一个窄带噪声过程$n(t)$。可将这个窄带过程表示为复基带形式$n(t) = \mathrm{Re}\{n_l(t)\, e^{j2\pi f_c t}\}$，其中$n_l(t)=n_I(t)+jn_Q(t)$为复低通高斯过程。$n_I(t)$和$n_Q(t)$为实低通高斯过程。可以证明，$n_I(t)$和$n_Q(t)$相互独立，它们每一个的均值都是零，功率谱密度都是是原噪声过程的两倍，为N_0。

随机过程的平稳性和广义平稳性是涉及概率空间的特性，我们也经常关心涉及时间平均的特性，这个特性用一些不同的*遍历性*（ergodic）来描述。随机过程$X(t)$的时间平均均值定义为：

$$\mu_X^{\mathrm{ta}} = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} X(t)\, \mathrm{d}t \tag{B-29}$$

若此值对$X(t)$的所有实现都相同，则称$X(t)$为*均值遍历的*（ergodic in the mean），也即对于$X(t)$的所有实现$x_i(t)$，$\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} x_i(t)\mathrm{d}t$ 都等于同一个常数μ_X^{ta}。类似地，随机过程$X(t)$是*n阶矩遍历的*（ergodic in n-th moment），若对于$X(t)$的所有实现，下式所定义的时间平均n阶矩是常数：

$$\mu_{X^n}^{\mathrm{ta}} = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} X^n(t)\, \mathrm{d}t \tag{B-30}$$

也可以针对时间平均自相关来谈论遍历性。时间平均自相关函数为：

$$A_X^{\mathrm{ta}}(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} X(t)X(t+\tau)\, \mathrm{d}t \tag{B-31}$$

如果对于$X(t)$所有的实现$x_i(t)$，$\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} x_i(t)x_i(t+\tau)\mathrm{d}t$ 都等于同一个值$A_X^{\mathrm{ta}}(\tau)$，则$X(t)$为*自相关遍历的*（ergodic in autocorrelation）。更高阶自相关遍历要求$X(t)$的所有实现的nm阶时间平均自相关函数为常数：

$$A_X^{\mathrm{ta}}(n,m,\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} X^n(t)X^m(t+\tau)\, \mathrm{d}t \tag{B-32}$$

如果一个过程的任意阶自相关都是遍历的，则称这个随机过程是*遍历的*。遍历过程要求对所有的n、i、j，其时间平均的n阶矩和ij阶自相关都是常数。这说明遍历过程所涉及的概率与时移无关，因此该过程必然是平稳的。也就是说，遍历过程一定是平稳过程。但平稳过程不一定是遍历过程。由于遍历过程是平稳过程，所以

$$
\begin{aligned}
\mu_X^{\text{ta}} &= E[\mu_X^{\text{ta}}] \\
&= E\left[\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} X(t)\,\mathrm{d}t\right] \\
&= \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} E[X(t)]\,\mathrm{d}t \\
&= \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \mu_X\,\mathrm{d}t = \mu_X
\end{aligned} \tag{B-33}
$$

585

因此，$X(t)$的时间平均的均值等于概率平均的均值。类似有

$$
\begin{aligned}
A_X^{\text{ta}}(\tau) &= E[A_X^{\text{ta}}(\tau)] \\
&= E\left[\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} X(t)X(t+\tau)\,\mathrm{d}t\right] \\
&= \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} E[X(t)(t+\tau)]\,\mathrm{d}t \\
&= \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} A_X(\tau)\,\mathrm{d}t = A_X(\tau)
\end{aligned} \tag{B-34}
$$

即时间平均的自相关函数等于概率平均的自相关函数。

B.4 高斯过程

通信系统中一般用高斯过程来描述噪声。对于任意的T及任意的函数$g(t)$，若

$$X_g = \int_0^T g(t)X(t)\,\mathrm{d}t \tag{B-35}$$

是高斯分布的随机变量，则称$X(t)$为高斯过程。接收机往往用积分器来进行信号检测，此时若接收机输入的噪声是高斯过程，则积分器输出的噪声随机变量是高斯分布的。X_g的均值为

$$E[X_g] = \int_0^T g(t)E[X(t)]\,\mathrm{d}t \tag{B-36}$$

方差为

$$\mathrm{Var}[X_g] = \int_0^T \int_0^T g(t)g(s)E[X(t)X(s)]\,\mathrm{d}t\,\mathrm{d}s - (E[X_g])^2 \tag{B-37}$$

如果$X(t)$为广义平稳，则上两式可简化为

$$E[X_g] = \int_0^T g(t)\mu_X\,\mathrm{d}t \tag{B-38}$$

$$\mathrm{Var}[X_g] = \int_0^T \int_0^T g(t)g(s)R_X(s-t)\,\mathrm{d}t\,\mathrm{d}s - (E[X_g])^2 \tag{B-39}$$

586

根据定义可以得到高斯随机过程的几个重要性质。首先，高斯过程通过线性时不变系统的输出也是一个高斯过程。其次，在式（B-35）中令$g(t)=\delta(t-t_i)$可知$X(t_i)$, $i=0,1,\cdots$是联合高斯随机变量。由于这些采样是高斯随机变量，故若样值不相关，则它们独立。此外，对于广义平稳高斯过程，式（B-35）中X_g的分布只取决于$X(t)$的均值和自相关。最后，注意随机过程的特性完全由任意采样时刻的样值的联合概率特性给定。对于高斯过程，这些样值是联合高斯的，联合分布由高斯过程的均值和自相关决定，因此高斯过程的概率特性完全由它的均值和自相关函数决定，所以广义平稳的高斯过程也是平稳过程。类似地，均值和自相关函数遍历的高斯过程也是遍历过程。

参考文献

[1] W. Feller, *An Introduction to Probability Theory and Its Applications,* vols. I and II, Wiley, New York, 1968 and 1971.

[2] R. M. Gray and L. D. Davisson, *Random Processes: A Mathematical Approach for Engineers,* Prentice-Hall, Englewood Cliffs, NJ, 1986.

[3] W. B. Davenport, Jr., and W. L. Root, *An Introduction to the Theory of Random Signals and Noise*, McGraw-Hill, New York, 1987.

[4] A. Leon-Garcia, *Probability and Random Processes for Electrical Engineering,* 2nd ed., Addison-Wesley, Reading, MA, 1994.

[5] P. Billingsley, *Probability and Measure,* 3rd ed., Wiley, New York, 1995.

[6] R. G. Gallager, *Discrete Stochastic Processes*, Kluwer, Dordrecht, 1996.

[7] H. Stark and J. W. Woods, *Probability and Random Processes with Applications to Signal Processing,* 3rd ed., Prentice-Hall, Englewood Cliffs, NJ, 2001.

[8] A. Papoulis and S. U. Pillai, *Probability, Random Variables and Stochastic Processes,* McGraw-Hill, New York, 2002.

587



附录C　矩阵的定义、运算和性质

本附录概述了书中所用到的矩阵方面的定义、运算和性质。关于更多矩阵方面的处理，以及本附录中给出的性质的证明，请参见文献[1~4]。

C.1　矩阵和向量

$N\times M$矩阵（matrix）$\boldsymbol{A}$是一个N行M列数字构成的方阵，记为：

$$\boldsymbol{A}=\begin{bmatrix} a_{11} & \cdots & a_{1M} \\ \vdots & \ddots & \vdots \\ a_{N1} & \cdots & a_{NM} \end{bmatrix} \tag{C-1}$$

$\boldsymbol{A}$的第ij个元素（即位于第i行第j列的元素）记为$\boldsymbol{A}_{ij}$。在式（C-1）中，$\boldsymbol{A}_{ij}=a_{ij}$。矩阵的元素也称为标量（scalar）以表明它是单个数字。对于$N\times M$矩阵，$N=M$称为方阵，$N>M$称为瘦阵，$N<M$称为胖阵。

方阵的对角元素指位于矩阵左上角到右下角的对角线上的元素，即$\boldsymbol{A}_{ij}$，$i=j$。$N\times N$方阵的迹（trace）为对角元素的和：$\mathrm{Tr}[\boldsymbol{A}]=\sum_{i=1}^{N}\boldsymbol{A}_{ii}$。若一个方阵的所有非对角元素都是零，即若$A_{ij}=0,\ i\neq j$，则称其为对角阵（diagonal matrix）。我们用$\mathrm{diag}[a_1,\cdots,a_N]$表示对角元素为$a_1,\cdots,a_N$的对角阵。$N\times N$单位阵$\boldsymbol{I}_N$是对角元素全为1的对角阵，即$\boldsymbol{I}_N=\mathrm{diag}[1,\cdots,1]$。在不至混淆的情况下，$\boldsymbol{I}_N$的下标$N$可以省略。

上三角（upper triangular）阵是这样一个方阵，其对角线之下的元素都为零，即$\boldsymbol{A}_{ij}=0,\ i>j$。下三角（lower triangular）阵则是对角线之上的元素都为零的方阵，即$\boldsymbol{A}_{ij}=0,\ i<j$。对角阵既是上三角阵，又是下三角阵。

只要维数没问题，矩阵的元素也可以是矩阵。例如，若$\boldsymbol{B}$是$N\times M_1$矩阵，$\boldsymbol{C}$是$N\times M_2$矩阵，则$\boldsymbol{A}=[\boldsymbol{B},\ \boldsymbol{C}]$形成了一个$N\times(M_1+M_1)$矩阵，也可以写成$\boldsymbol{A}=[\boldsymbol{B}|\boldsymbol{C}]$，其第$i$行为$[\boldsymbol{A}_{i1}\cdots\boldsymbol{A}_{i(M_1+M_2)}]=[\boldsymbol{B}_{i1}\cdots\boldsymbol{B}_{iM_1}\boldsymbol{C}_{i1}\cdots\boldsymbol{C}_{iM_2}]$。若另有$K\times L_1$矩阵$\boldsymbol{D}$和$K\times L_2$矩阵$\boldsymbol{E}$，且$M_1+M_2=L_1+L_2$，则可以组成一个$(N+K)\times(M_1+M_2)$矩阵：

$$\boldsymbol{A}=\begin{bmatrix} \boldsymbol{B} & \boldsymbol{C} \\ \boldsymbol{D} & \boldsymbol{E} \end{bmatrix} \tag{C-2}$$

矩阵$\boldsymbol{B}$、$\boldsymbol{C}$、$\boldsymbol{D}$和$\boldsymbol{E}$称为$\boldsymbol{A}$的子阵（submatrix）。只要大小合适，一个矩阵可以由任意多个子阵组成。也可以通过删除$\boldsymbol{A}$的某些行或列得到子阵$\boldsymbol{A}'$。

只有一个列的矩阵（即$M=1$）称为列向量（column vector）或向量（vector）。向量的行数称为其维数。例如下式的$\boldsymbol{x}$就是一个N维向量：

$$\boldsymbol{X}=\begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix} \tag{C-3}$$

向量$\boldsymbol{x}$的第i个元素记为x_i。称所有元素均为1的N维向量为全幺向量，记为$\boldsymbol{I}_N$。只有一个

元素为1，其余元素为零的向量称为单位向量。第i个单位向量e^i有$e_i^i=1$，$e_i^j=0$，$j\neq i$。只有一行的矩阵（即$N=1$）称为**行向量**（row vector）。行向量的列数是其维数。M维行矢量为$x=[x_1\cdots x_M]$，第i个元素$x_i=x_i$。N维行向量或者列向量的**欧氏范数**（Euclidean norm），也称**范数**（norm），定义为：

$$\|X\| = \sqrt{\sum_{i=1}^{N}|X_i|^2} \tag{C-4}$$

C.2　矩阵和向量的运算

$N\times M$矩阵A的**转置**（transpose）A^T是$M\times N$矩阵，定义为$A^T_{ij}=A_{ji}$：

$$A^T = \begin{bmatrix} a_{11} & \cdots & a_{1M} \\ \vdots & \ddots & \vdots \\ a_{N1} & \cdots & a_{NM} \end{bmatrix}^T = \begin{bmatrix} a_{11} & \cdots & a_{N1} \\ \vdots & \ddots & \vdots \\ a_{1M} & \cdots & a_{NM} \end{bmatrix} \tag{C-5}$$

A^T也就是调换了A的行和列，A的第i行是A^T的第i列。行向量$x=[x_1\cdots x_N]$的转置是相同元素组成的列向量：

$$X^T = [X_1 \ \ldots \ X_N]^T = \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix} \tag{C-6}$$

因此我们经常把列矢量写成$x=[x_1\cdots x_N]^T$的形式。类似地，N维（列）向量x的转置是行向量$[x_1\cdots x_N]$。无论x是行向量还是列向量，总有$(x^T)^T=x$。

589

矩阵A的**复共轭**（complex conjugate）A^*是将A的每个元素取复共轭：

$$A^* = \begin{bmatrix} a_{11} & \cdots & a_{1M} \\ \vdots & \ddots & \vdots \\ a_{N1} & \cdots & a_{NM} \end{bmatrix}^* = \begin{bmatrix} a^*_{11} & \cdots & a^*_{1M} \\ \vdots & \ddots & \vdots \\ a^*_{N1} & \cdots & a^*_{NM} \end{bmatrix} \tag{C-7}$$

矩阵A的**厄密共轭**（Hermitian）A^H是其共轭转置：$A^H=(A^*)^T$。两次厄密运算的结果是原矩阵：$(A^H)^H=A$，故A也是A^H的厄密共轭。若满足$A=A^H$，称方阵A为厄密矩阵。若满足$A^HA=AA^H$，称方阵A为**正规矩阵**（normal matrix）。厄密矩阵同时也是正规矩阵，因为$A^H=A$，所以$A^HA=AA^H$。也可以对向量做复共轭和厄密运算。列向量或者行向量x的复共轭x^*是对x的每个元素取复共轭。向量x的厄密向量x^H是其共轭转置：$x^H=(x^*)^T$。

两个$N\times M$矩阵可以相加，结果是一个新的$N\times M$矩阵，加法是逐元素相加。也就是说，两个$N\times M$矩阵A和B的和$C=A+B$是$N\times M$矩阵，其第ij个元素是$C_{ij}=A_{ij}+B_{ij}$。因为是逐元素相加，所以加法的交换律和结合律仍然成立，即$A+B=B+A$，$(A+B)+C=A+(B+C)$。矩阵和的转置等于矩阵转置的和：$(A+B)^T=A^T+B^T$。矩阵的减法与此类似，对于两个$N\times M$矩阵A和B，$C=A-B$是$N\times M$矩阵，其第ij个元素是$C_{ij}=A_{ij}-B_{ij}$。行向量或列向量是矩阵的特例，也可以按矩阵加法的定义进行相加，例如N维向量x、y相加得$z=x+y$，其第i个元素为$z_i=x_i+y_i$。行向量类似。但一个维数$N>1$的行向量不能同一个N维列向量相加，

因为它们是大小不同的矩阵（行向量是$1 \times N$，列向量是$N \times 1$）。N维向量$\boldsymbol{x}$、$\boldsymbol{y}$的线性组合$z = c\boldsymbol{x}+d\boldsymbol{y}$是一个新的$N$维向量，其第$i$个元素是$z_i = cx_i+dy_i$，$c$和$d$是任意的标量。

可以对矩阵乘以任意标量，其结果是矩阵的每一个元素都乘以这个标量。如矩阵$\boldsymbol{A}$乘以标量k的结果$k\boldsymbol{A}$为：

$$k\boldsymbol{A} = k\begin{bmatrix} a_{11} & \cdots & a_{1M} \\ \vdots & \ddots & \vdots \\ a_{N1} & \cdots & a_{NM} \end{bmatrix} = \begin{bmatrix} ka_{11} & \cdots & ka_{1M} \\ \vdots & \ddots & \vdots \\ ka_{N1} & \cdots & ka_{NM} \end{bmatrix} \tag{C-8}$$

行向量$\boldsymbol{x}$乘以标量k的结果是$k\boldsymbol{x}=[kx_1 \cdots kx_N]$，列向量$\boldsymbol{x}$乘以标量$k$的结果是$k\boldsymbol{x}=[kx_1 \cdots kx_N]^T$。 590

维数相容的两个矩阵可以相乘，具体要求是第一个矩阵的列数应等于第二个矩阵的行数。若$\boldsymbol{A}$是$N \times M$矩阵，$\boldsymbol{B}$是$M \times L$矩阵，则$\boldsymbol{C}=\boldsymbol{A}\boldsymbol{B}$是$N \times L$矩阵，其第$ij$个元素是$C_{ij} = \sum_{k=1}^{M} A_{ik}B_{kj}$。矩阵乘法一般不满足交换率（即一般$\boldsymbol{AB} \neq \boldsymbol{BA}$）。实际上，对于$N \times M$矩阵$\boldsymbol{A}$和$M \times L$矩阵$\boldsymbol{B}$，积$\boldsymbol{BA}$仅当$L=N$时存在，此时$\boldsymbol{BA}$是$M \times M$矩阵，它与$N \times L$的矩阵$\boldsymbol{AB}$可能大小都不一样。即便$M=L=N$，$\boldsymbol{AB}$和$\boldsymbol{BA}$这两个矩阵虽然大小相同，但内容不一定相等。方阵$\boldsymbol{A}$可以自己乘自己，定义$\boldsymbol{A}^2=\boldsymbol{AA}$，$\boldsymbol{A}^k=\boldsymbol{A}\cdots\boldsymbol{A}$为$k$个$\boldsymbol{A}$的连乘，由此可得$\boldsymbol{A}^k\boldsymbol{A}^l = \boldsymbol{A}^{k+l}$。任意矩阵与维数相容的单位阵相乘的结果还是原矩阵，即若$\boldsymbol{A}$为$N \times M$矩阵，则$\boldsymbol{I}_N\boldsymbol{A} = \boldsymbol{A}\boldsymbol{I}_M = \boldsymbol{A}$。两矩阵积的转置等于每个矩阵转置后以相反的顺序相乘：$(\boldsymbol{AB})^T = \boldsymbol{B}^T\boldsymbol{A}^T$。$N \times M$矩阵$\boldsymbol{A}$与它的$M \times N$厄密阵$\boldsymbol{A}^H$的积为方阵，$\boldsymbol{AA}^H$为$N \times N$方阵，$\boldsymbol{A}^H\boldsymbol{A}$为$M \times M$方阵。矩阵$\boldsymbol{A}$的Frobenius范数定义为$\|\boldsymbol{A}\|_F = \sqrt{\mathrm{Tr}[\boldsymbol{AA}^H]} = \sqrt{\mathrm{Tr}[\boldsymbol{A}^H\boldsymbol{A}]} = \sqrt{\sum_{i=1}^{N}\sum_{j=1}^{M}|A_{ij}|^2}$。只要矩阵的维数相容，矩阵乘法满足结合律：$(\boldsymbol{AB})\boldsymbol{C} = \boldsymbol{A}(\boldsymbol{BC})$，故通常省略括号。矩阵相乘也满足分配律：$\boldsymbol{A}(\boldsymbol{B}+\boldsymbol{C})=\boldsymbol{AB}+\boldsymbol{AC}$, $(\boldsymbol{A}+\boldsymbol{B})\boldsymbol{C}=\boldsymbol{AC}+\boldsymbol{BC}$。

M维向量可以与M列的矩阵相乘。若$\boldsymbol{A}$是一个$N \times M$矩阵而$\boldsymbol{x}$是一个M维向量（即$M \times 1$矩阵），则它们的积是N维向量$\boldsymbol{y} = \boldsymbol{Ax}$，第$i$个元素为$y_i = \sum_{k=1}^{M} A_{ik}x_k$。注意必须是矩阵左乘向量，因为$\boldsymbol{xA}$存在维数不相容的问题。但若$\boldsymbol{x}$是$N$维行向量，则对于$N \times M$矩阵$\boldsymbol{A}$，$\boldsymbol{xA}$维数相容，结果是一个$M$维行向量，第$i$个元素是$y_i = \sum_{k=1}^{M} x_k A_{ki}$。$N$维行向量和$N$维列向量相乘的结果是一个标量$z = \boldsymbol{xy} = \sum_{i=1}^{N} x_i y_i$。注意到$N$维向量的转置是$N$维行向量，因此定义两个$N$维向量的内积为$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{x}^T\boldsymbol{y} = \sum_{i=1}^{N} x_i y_i$。

对于矩阵$\boldsymbol{A}$的一些行组成的子集，若其中任一行都不是其他行的线性组合，则这个子集是*线性无关的*（linearly independent）。类似地，矩阵$\boldsymbol{A}$的一些列组成了一个线性无关子集，若其中的任一列都不是其他列的线性组合。矩阵$\boldsymbol{A}$的*秩*（rank）R_A等于矩阵$\boldsymbol{A}$的线性无关行构成的最大子集的行数，可证明它也等于矩阵$\boldsymbol{A}$的线性无关列构成的最大子集的列数。这表明$N \times M$矩阵的秩不能超过$\min[N, M]$。若$R_A=\min[N, M]$，则称$N \times M$矩阵$\boldsymbol{A}$是满秩的。

2×2矩阵$\boldsymbol{A}$的*行列式*（determinant）定义为$\det[\boldsymbol{A}]=A_{11}A_{22}-A_{21}A_{12}$。对于$N \times N$（$N>2$）矩阵$\boldsymbol{A}$，$\det[\boldsymbol{A}]$以迭代方式定义为：

$$\det[A] = \sum_{i=1}^{N} A_{ij} c_{ij} \tag{C-9}$$

591 其中j是$1 \leqslant j \leqslant N$之间的任意整数，$c_{ij}$是矩阵元素$A_{ij}$的余子式（cofactor），定义为：

$$c_{ij} = (-1)^{i+j} \det[A'] \tag{C-10}$$

其中A'是从A中删除第i行第j列后得到的子阵。

对于$N \times N$方阵A，若存在另外一个$N \times N$方阵B使得$BA=I_N$，则称A是可逆的（invertible），或者非奇异的（nonsingular），称B为A的逆（inverse），记为A^{-1}。于是$A^{-1}A = I_N$。对于这样定义的A^{-1}，也有$AA^{-1}= I_N$。只有方阵是可逆的，并且逆矩阵和原矩阵大小相同。若满足$UU^H = I$，这表明$U^H = U^{-1}$，因而$U^HU=I$，这时称可逆方阵U为酉（unitary）阵。并非所有的方阵都是可逆的，若其不可逆，则称其为奇异的（singular）或者不可逆的（noninvertible）方阵。逆矩阵的逆是原矩阵：$(A^{-1})^{-1}=A$。矩阵积的逆为矩阵的逆以相反顺序的积：$(AB)^{-1}=B^{-1}A^{-1}$。逆的k次幂为$A^{-k}=(A^{-1})^k$。

对角阵$D=\mathrm{diag}[d_1, \cdots, d_N]$当所有$d_i \neq 0(i=1, \cdots,N)$时，其逆存在且为$D^{-1}=\mathrm{diag}[1/d_1, \cdots, 1/d_N]$。对于一般的$2 \times 2$矩阵$A$，当$\det[A] \neq 0$时逆存在，为：

$$A^{-1} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}^{-1} = \frac{1}{\det[A]} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix} \tag{C-11}$$

对于大于2×2的可逆矩阵有更复杂的求逆公式，一般可用数学软件包来求解。

矩阵的逆通常用于求解线性方程组。考虑一个矩阵形式表示的线性方程组：

$$y = Ax \tag{C-12}$$

如果矩阵A可逆，则给定y时，这个方程组有唯一解$x = A^{-1}y$。

C.3 矩阵分解

给定方阵A，若对于标量λ存在一个非零向量x使得$Ax=\lambda x$，则称此标量为矩阵A的特征值（eigenvalue）。这个向量x称为矩阵A对应于λ的特征向量（eigenvector）。矩阵A的所有特征值是其特征方程（characteristic equation）$\det[A-\lambda I] = 0$的解。$\det[A-\lambda I]$是关于λ的多项式，称为矩阵A的特征多项式（characteristic polynomial），因此A的所有特征值都是其特征多项式的根。$N \times N$矩阵的特征多项式，若其形式为$\det[A-\lambda I] = (-1)^N(\lambda-r_1)\cdots(\lambda-r_N)$，则有$N$个各不相同的根$r_1, \cdots, r_N(r_i \neq r_j)$。当特征多项式包含$(\lambda-r_i)^k$，$k>1$时，称$r_i$为$k$重根。例如：若$\det[A-\lambda I]=-(\lambda-r_1)^2(\lambda-r_2)^3$，则$r_1$是2重根，$r_2$为3重根。$N \times N$矩阵有$N$个特征值$\lambda_1,\cdots,\lambda_N$，其中有些可能因为重根而相同。可以证明，矩阵的行

592 列式等于其所有特征值的积（求积时k重的特征值r_i按r_i^k计算）。

厄密矩阵的特征向量可能是复数的，但特征值总是实数。此外，$N \times N$的正规矩阵A可以写成下面的形式：

$$A = U\Lambda U^H \tag{C-13}$$

其中U是酉阵，其列是A的特征向量。$\Lambda=\mathrm{diag}[\lambda_1,\cdots,\lambda_K,0,\cdots,0]$是$N\times N$对角阵，前$K$个对角元素是$A$的非零特征值。对于厄密矩阵，式（C-13）中的$\Lambda$只有实元素。称矩阵$A$为*正定的*（positive definite），若对任意的非零矢量x有$x^H Ax>0$。厄密矩阵是正定的，当且仅当其所有特征值都是正的。类似地，称矩阵A为*半正定的*（positive semidefinite）或者*非负定的*（nonnegative definite），若对任意的非零矢量x有$x^H Ax\geqslant 0$。厄密矩阵为非负定的，当且仅当其所有特征值都是非负的。

若$N\times M$矩阵A的秩为R_A，则存在一个$N\times M$矩阵Σ，一个$N\times N$酉阵U，和一个$M\times M$酉阵V，使得：

$$A=U\Sigma V^H \tag{C-14}$$

称V的列为A的*右奇异向量*（right singular vector），U的列为A的*左奇异向量*（left singular vector）。矩阵Σ的非对角元素均为零，当$N\geqslant M$时形为：

$$\Sigma_{N\times M}=\begin{bmatrix}\sigma_1 & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & \sigma_M\\ 0 & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & 0\end{bmatrix} \tag{C-15}$$

当$N<M$时形为：

$$\Sigma_{N\times M}=\begin{bmatrix}\sigma_1 & \cdots & 0 & 0 & \cdots & 0\\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots\\ 0 & \cdots & \sigma_N & 0 & \cdots & 0\end{bmatrix} \tag{C-16}$$

这里$\sigma_i=\sqrt{\lambda_i}$，$\lambda_i$是$AA^H$的第$i$个特征值。$\sigma_i$称为$A$的*奇异值*（singular value），其中有R_A个非零值，R_A是A的秩。式（C-14）的分解称为矩阵A的*奇异值分解*（singular value decomposition，SVD）。矩阵的奇异值总是非负的。

593

令A是一个$N\times M$矩阵，记其第i列为A_i。把每个列当作一个子矩阵，可以将A写成$A=[A_1A_2\cdots A_M]$。矩阵A的向量化vec(A)是把A的列A_i自上而下排列得到的一个NM维向量：

$$\mathrm{vec}(A)=\begin{bmatrix}A_1\\ \vdots\\ A_M\end{bmatrix}=[A_{11}\ A_{21}\ \cdots\ A_{N1}\ A_{12}\ \cdots\ A_{N2}\ \cdots\ A_{1M}\ \cdots\ A_{NM}]^T \tag{C-17}$$

令A为$N\times M$矩阵，B为$L\times K$矩阵。矩阵A和矩阵B的Kronecker积$A\otimes B$是一个$NL\times MK$矩阵，定义为：

$$A\otimes B=\begin{bmatrix}A_{11}B & \cdots & A_{1M}B\\ \vdots & \ddots & \vdots\\ A_{N1}B & \cdots & A_{NM}B\end{bmatrix} \tag{C-18}$$

参考文献

[1] B. Nobel and J. W. Daniel, *Applied Linear Algebra*, Prentice-Hall, Englewood Cliffs, NJ, 1977.
[2] G. Strang, *Linear Algebra and Its Applications*, 2nd ed., Academic Press, New York, 1980.
[3] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, 1985.
594 [4] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*, Cambridge University Press, 1991.



附录D　无线标准摘要

本章摘要描述了当今运行的最普遍的两类无线通信系统的技术细节，它们是蜂窝电话和无线局域网。本章还摘要描述了三种短距离的无线网络标准，它们已经获得了广泛的应用。无线标准的详情请参考文献[1~6]。

D.1　蜂窝电话标准

D.1.1　第一代模拟系统

这一节摘要描述蜂窝电话标准。我们从第一代（1G）模拟蜂窝电话标准开始，其主要特性参考表D-1。基于这些标准的系统在20世纪80年代被广泛采用。尽管现在很多这样的系统已经被数字蜂窝系统取代，在全世界的许多地方模拟系统还在使用。最知名的模拟系统标准为Advanced Mobile Phone System（AMPS），它是20世纪70年代由贝尔实验室开发，1983年在美国投入商用。美国采用以后，许多其他国家也接受了它。AMPS有一个窄带的版本，称为窄带AMPS（N-AMPS），其话音信道带宽是普通AMPS的三分之一。1979年，日本开发了它的第一个商用蜂窝电话系统，基于AMPS的NTT标准（MCS-L1）。欧洲也发展了类似AMPS的标准，称为Total Access Communication System（TACS）。与AMPS相比，TACS采用更高的频率和更低的信道带宽。它在英国、欧洲国家以及欧洲以外的国家被采用。在英国，TACS的频率范围被扩展，以获得更多的信道，产生了叫做ETACS的变种版本。1989年，称为JTACS的TACS的变种在日本的大都市地区被采用，用以获得比NTT系统更高的容量。JTACS运行在比TACS和ETACS稍微高一些的频率上，而且还有一种提高频带利用率的版本，称为NTACS，其语音信道的带宽为JTACS信道带宽的一半。除TACS以外，欧洲国家还有频率上不兼容的模拟蜂窝电话标准，包括斯堪迪纳维亚的Nordic Mobile Telephone（NMT）标准，法国的Radiocom 2000（RC2000）标准，德国和波兰的C-450标准。不兼容使得用一个模拟电话漫游欧洲国家成为一件不可能的事情，这引发了全欧洲统一的蜂窝电话标准和频率的需求。

595

表D-1　第一代模拟蜂窝电话标准

参　数	AMPS	TACS	NMT（450/900）	NTT	C-450	RC2000
上行链路频率（MHz）	824-849	890-915	453-458/890-915	925-940[a]	450-455.74	414.8-418[b]
下行链路频率（MHz）	869-894	935-960	463-468/935-960	870-885	460-465.74	424.8-428
调制方式	FM	FM	FM	FM	FM	FM
信道宽度（kHz）	30	25	25/12.5	25	10	12.5
信道数量	832	1000	180/1999	600	573	256
多址方式	FDMA	FDMA	FDMA	FDMA	FDMA	FDMA

a. NTT也用于900MHz左右的频率范围。

b. RC2000也用于200MHz左右的频率范围。

D.1.2　第二代数字系统

下面我们讨论第二代（2G）数字蜂窝电话系统标准，其主要特征归纳在表D-2中。这些系统基本上在90年代投入使用。由于第一代系统的不兼容性，1982年，名为Group Special Mobile（GSM）的组织被建立起来，发展了一个全欧洲的统一的数字蜂窝标准。在欧洲，TACS的900MHz的频段分配给了GSM以推动国家之间的漫游。1989年，GSM规范定稿，1991年，系统开通。在1992年以前，GSM系统还不是很有效。GSM标准使用TDMA和慢跳频对抗越区干扰，使用带交织的卷积码和奇偶校验码来检错和纠错。这个标准也包括了一个均衡器，用以补偿频率选择性衰落。世界上超过66%的蜂窝电话使用GSM标准，超过470个运营商在172个国家支持着数以十亿计的用户。随着GSM标准越来越全球化，这个缩写的意思变成了Global System for Mobile Communications。

表D-2　第二代数字蜂窝电话标准

参　数	GSM	IS-136	IS-95（cdmaOne）	PDC
上行链路频率（MHz）	890-915	824-849	824-849	940-956,1429-1453
下行链路频率（MHz）	935-960	869-894	869-894	810-826,1477-1501
载波间隔（kHz）	200	30	1250	25
信道数量	1000	832[a]	~2500	1600[a]
调制方式	GMSK	π/4-D-QPSK	BPSK/QPSK	π/4-D-QPSK
压缩话音速率（kbit/s）	13	7.95	1.2-9.6（可变的）	6.7
信道数据速率（kbit/s）	270.833	48.6	（1.2288Mchips/s）	42
数据码率	1/2	1/2	1/2（DL），1/3（UL）	1/2
ISI消除/均衡	均衡器	均衡器	瑞克接收，软切换	均衡器
多址接入	TDMA/慢调频	TDMA	CDMA	TDMA

a. 每个通道3个用户。

虽然在开发2G数字系统的过程中，欧洲走在了前面，但是美国落后得也不远。1992年，IS-54数字蜂窝标准在美国定稿，1994年开始商用。这个标准采用了与AMPS同样的30kHz的信道宽度，用以促进无线运营商从模拟向数字转移，采用TDMA多址接入方式，用以改进切换和控制信令。IS-54标准，也称为North American Digital Cellular标准。随着时间的推移，IS-54标准进行了改进，这些改进被封装在IS-136标准中，IS-136标准包含了原来的标准。类似于GSM标准，IS-136标准采用了奇偶校验码，卷积码，交织，均衡。 596

另外一个竞争市场的2G标准是基于CDMA的，它由高通公司在20世纪90年代初期提出。这个称为IS-95或者IS-95a的标准，在1993年定稿，在1995年用cdmaOne的名字投入商用。跟IS-136一样，IS-95设计成与AMPS兼容，以便这两个系统在同一个频带内共存。在CDMA系统中，用户彼此叠加，接收机通过扩频码区分出不同的用户。这样，信道数据速率就不像TDMA那样仅仅赋给一个用户。信道码片速率为1.2288M chips/s，扩频因子为128。IS-95中，上行链路和下行链路的扩频过程是不同的，尽管它们都通过扩频调制和编码完成。对于下行链路，数据首先进行码率为1/2的卷积编码和交织，然后

由64个正交扩频序列（Walsh函数）之中的一个调制。之后把一个对于每个小区唯一的同步扰码序列叠加在Walsh函数上，以减小小区间的干扰，扰码要求基站之间同步。对于上行链路，使用码率为1/3的带交织的卷积码，用正交Walsh函数调制，并用非正交的用户/基站代码进行调制。IS-95标准包括一个用于错误检测的奇偶校验码，以及避免远近效应的反向链路的功率控制，还定义了一个三路RAKE接收机，来利用分集和补偿ISI。软切换（SHO）作为一种基站分集也包含在这个标准中，在切换期间，移动台保持与新旧两个基站的连接，并把它们的信号合成起来。在蜂窝系统中，CDMA相对TDMA有一些优点，包括不需要频率规划，SHO能力，使用语音激活增加容量，系统容纳的用户数量没有硬性的限制等。在20世纪90年代早期，有很多关于IS-136和IS-95标准优缺点的争论，声称IS-95可以达到AMPS 20倍的容量，而IS-136只能达到AMPS 3倍的容量。最后证实，相对AMPS，两个系统具有大致相同的容量增长。

在日本，2G数字蜂窝标准称为Personal Digital Cellular（PDC）标准，它在1991年建立，1994年被采用。它类似于IS-136，但是采用25kHz话音信道以兼容日本的模拟系统。这种系统运行在800MHz~900MHz和1400MHz频带上。每个信道可容纳3个用户，当采用高压缩率时每个信道能容纳6个用户。

597

D.1.3 2G系统的演化

20世纪90年代后期，2G系统在两个方向上演进：系统转向更高的频率，这是由于在欧洲和美国，有更多的蜂窝带宽可以获得。同时，系统被改进以支持话音业务以外的数据业务。1994年FCC开始拍卖位于1.9GHz的Personal Communication System（PCS）的蜂窝系统频带。购买这个频段内的频谱的运营商可以采用任何标准。由于不同的运营商采用不同的标准，于是在美国的不同地区，1900MHz频段上GSM，IS-136，IS-95都被采用，造成了单个电话在国内漫游的困难。事实上，许多早期的蜂窝电话，包含了模拟的AMPS模式，用以防备数字系统不能获得服务的情况。运行在PCS频段的GSM系统称为PCS1900系统，而迁移到PCS频段的IS-136和IS-95标准依旧使用相同的名字。欧洲在1.8GHz频段分配附加的蜂窝电话频谱，这个频段的标准，称为GSM1800或者DCS1800，其采用GSM作为核心标准，做了一些改进，允许宏蜂窝和微蜂窝的覆盖。注意，第二代无绳电话，如DECT，Personal Access Communications System（PACS），以及Personal Handyphone System（PHS）也运行在1900MHz频段，但是这些系统大多是在建筑物内提供专用交换服务（PBX）。

一旦数字蜂窝系统可以使用，运营商就会引入话音业务之外的数据业务。增加数据能力的2G系统，有时称为2.5G系统。为了支持数据业务而对2G系统做的增强，列在表D-3中。GSM系统可以有几个不同的提供数据服务的升级路径。最简单的，称为High Speed Circuit Switched Data（HSCSD），允许最多4个连续时隙分配给一个用户，这样可以提供最高达到57.4kbit/s的传输速率。对于数据来说，电路交换效率很低。因而，更复杂的增强是在电路交换的话音业务上提供分组交换的数据业务。这种增强称为General Packet Radio Service（GPRS）。对于GPRS来说，当GSM帧的8个时隙分配给一个用户的



时候，最大可能的数据速率为171.2kbit/s。GPRS的数据速率通过变速率调制和编码被进一步增强，这称为Enhanced Data rates for GSM Evolution（EDGE）。EDGE提供最高达到384kbit/s的数据速率，其每个时隙的比特速率为48kbit/s~69.2kbit/s。GPRS和EDGE与IS-136和GSM兼容，这样为这两个系统提供了一个会聚在一起的升级路径。

598

IS-95标准通过分配多个正交Walsh函数给各个用户来提供数据业务。最多可以分配8个函数，达到最大115.2kbit/s的数据速率，实际上大约可以达到64kbit/s。这种演进称为IS-95b标准。

表D-3　支持2.5G数据能力的增强项目

2G标准	2.5G标准	函　数	数据速率：最大/实际
GSM	HSCSD	技术	57.6/14.4-57.6 kbit/s
GSM/IS-136	GPRS	合并时隙	140.8/56 kbit/s
	EDGE	合并时隙，可变调制，	384/200 kbit/s
IS-95	IS-95b	合并Walsh	115/64 kbit/s

D.1.4　第三代系统

2G系统标准和频率的零乱，导致国际电信联盟（ITU）在20世纪90年代制定了一个第三代数字蜂窝电话（3G）的计划，采用全球统一的频段和标准。这个标准命名为International Mobile Telephone 2000 （IMT-2000）标准，并要求在2000年的时限内展示系统。除了话音业务以外，IMT-2000为苛刻的应用提供Mbit/s级的数据速率，这些应用包括宽带互联网接入，交互游戏，以及高质量的音频和视频娱乐等。单一标准的共识并没有实现，多数国家支持两个竞争的标准之中的一个，CDMA2000（后向兼容CDMAOne）被Thrid Generation Partnership Project 2 （3GPP2）支持，wideband CDMA （W-CDMA，后向兼容GSM和IS-136）被Thrid Generation Partnership Project 1 （3GPP1）支持。这两个3G标准的主要特征归纳在表D-4中。两个标准都采用了功率控制的CDMA和RAKE接收机，但是码片速率和其他规格的细节不同。特别地，CDMA2000和W-CDMA不是兼容的标准，因而，一个移动台要在两种系统里都能运行，则必须是双模式的。第三个3G标准，TD-SCDMA，在中国正在被考虑之中，但是未必会在其他地方采用。TD-SCDMA与其他3G标准的关键不同是，对于上行下行链路，它使用TDD而不是FDD。

599

基于CDMAOne的CDMA2000标准提供了一条向3G演进的路径。CDMA2000的核心标准为CDMA2000 1X或者CDMA2000 1XRTT，表示无线传输技术（RTT）在一对1.25MHz的无线信道中运行，并且因此与CDMAOne后向兼容。CDMA2000 1X系统加倍了CDMAOne系统的话音容量，并提供高速数据业务，其规划峰值速率约300kbit/s，实际速率约为144kbit/s。其核心技术经过两次演进，以提供超过1Mbit/s的高数据速率（HDR业务）：这些演进称为CDMA2000 1XEV。演进的第一阶段，CDMA2000 1XEV-DO （Data Only），使用一个单独的1.25MHz专用高速数据信道支持高达3Mbit/s的下行链路和高达1.8Mbit/s的上行链路，其平均组合速率为2.4Mbit/s。演进的第二阶段，

CDMA2000 1XEV-DV （Data and Voice），规划用同样的无线信道支持高达4.8Mbit/s的数据速率，也支持1X话音用户，1XRTT数据用户，以及1XEV-DO数据用户。CDMA2000的另外一种被提议的增强方式为组合3个1.25MHz信道形成一个3.75MHz信道，称为CDMA2000 3X，其详细的规范还在开发之中。

表D-4　第三代数字蜂窝电话标准

3G标准	CDMA2000				W-CDMA		
子类	1X	1XEV-DO	1XEV-DV	3X	UMTS	FOMA	J-Phone
信道带宽（MHz）	1.25	1.25		3.75	5		
码片速率（Mchips/s）	1.2288			3.6864	3.84		
峰值数据速率（Mbit/s）	0.144	2.4	4.8	5~8	2.4[a]（HSDPA 8-10）		
调制	QPSK（下行）,BPSK（上行）						
编码	卷积码（低速）,Turbo码（高速）						
功率控制	800Hz				1500Hz		

a. 8Mbit/s~10Mbit/s具有HSDPA。

W-CDMA是CDMA2000的主要竞争标准。它已经被选为GSM的接任者，在本文中称为Universal Mobile Telecommunications System （UMTS）。W-CDMA也用于日本的FOMA和J-Phone 3G系统中。这些不同的系统共享W-CDMA链路级协议（空中接口），但是在系统的其他方面有不同的协议，例如路由和话音压缩。W-CDMA支持2.4Mbit/s 的峰值速率，典型速率预计为384kbit/s。相对于CDMA2000的1.25MHz 的信道，W-CDMA采用5MHz信道。W-CDMA的一种增强版本，称为High Speed Data Packet Access（HSDPA），提供大约9Mbit/s的数据速率，这个版本可能成为第四代系统的先驱。

D.2　无线局域网

无线局域网遵循IEEE802.11协议族。这个协议族的主要特性总结在表D-5中。802.11的原始版本，发表于1997年，在不需要许可证的2.4GHz频段使用83.5MHz带宽。它指定使用FHSS或者DSSS的PSK调制，CSMA/CA用于随机接入。原始的标准在1999年发展成了802.11b标准，运行在同样的2.4GHz频带上，但是只采用DSSS。这个标准采用变速率调制和编码，BPSK或者QPSK调制，信道编码采用Barker序列或者Cmoplementart Code Keying（CCK）。其最大信道速率可以达到11Mbit/s，最大用户数据速率大约为1.6Mbit/s，传输距离大约为100m。802.11b的网络结构为星形或者对点，其中对点结构不是典型的应用结构。这个标准被广泛采用，制造商把802.11b网卡集成到很多笔记本电脑中。

600

802.11a标准作为802.11的扩展，在1999年定稿，改进后的速率达到802.11b的数据速率。802.11a占用5GHz U-NII频段的300MHz频带。实际上，这300MHz频带别分成三个100MHz的子频带。低频带从5.15到5.25GHz，中频带从5.25到5.35GHz，高频带从

5.725到5.825GHz。信道之间有20MHz的间隔，而在低频带和中频带的边缘，有30MHz的间隔。三个最大传输功率定义为：低频带40mW，中频带200mW，高频带800mW。这个限定意味着，低频带更适合室内应用，中频带适合室内和户外应用，高频带适合户外应用。变速率调制和编码在每个信道上都被采用。调制在BPSK，QPSK，16QAM，64QAM中变化，卷积码率在1/2，2/3，3/4中变化。这使得每个信道的最大数据速率为54Mbit/s。对于室内系统，相对于802.11b，5GHz的载波以及低频带的功率限制降低了802.11a的覆盖范围，也使得它穿透墙壁和其他障碍物的更为困难。802.11a使用正交频分复用（OFDM）多址接入，而不是用FHSS或者DSSS。在这个意义上，它已经脱离了原始的802.11标准。

表D-5 802.11无线局域网链路层标准

参　数	802.11	802.11a	802.11b	802.11g
带宽（MHz）	83.5	300	83.5	83.5
频率范围（GHz）	2.4-2.4835	5.15-5.25（低） 5.25-5.35（中） 5.725-5.825（低）	2.4-2.4835	2.4-2.4835
信道数量	3	12（每个子信道4个）	3	3
调制	BPSK,QPSK, DSSS,FHSS	BPSK,QPSK, MQAM,OFDM	BPSK,QPSK,DSSS	BPSK,QPSK, MQAM,OFDM
编码	未定义	卷积码 （码率1/2,2/3,3/4）	Barker,CCK	卷积码 （码率1/2,2/3,3/4）
最高速率（MHz）	2	54	11	54
覆盖范围		27~30（低频带）	75~100	30
随机接入		CSMA/CA		

802.11g标准，定稿于2003年，试图融合802.11a和802.11b，它运行在2.5GHz或者更高频段，数据速率高达54Mbit/s。这个标准后向兼容802.11b，因此802.11g接入点可以与802.11b网卡一起工作，反之亦然。802.11g采用802.11a的OFDM技术，调制以及编码体系。接入点和无线网卡都支持三种标准以避免兼容问题。802.11a/b/g协议族统称Wi-Fi。在美国以外的其他国家，扩展这些标准的频率范围，产生了802.11d标准。还有一些802.11家族的其他标准正在开发之中。它们在表D-6中做了归纳。

表D-6 IEEE 802.11 进行中的工作

标　准	内　容
802.11e	在MAC层提供QoS
802.11f	不同运营商接入点之间的漫游协议
802.11h	为802.11a加入频率和功率控制特性以更适用于欧洲
802.11i	增强安全和鉴权机制
802.11j	修改802.11a的链路层以满足日本的要求
802.11k	为无线和网络测量提供一个到高层的接口，可以用于无线资源管理
802.11m	标准的维护（技术/编辑更正）
802.11n	MIMO链路增强，以使吞吐率更高

802.11标准以及蜂窝系统的一个潜在竞争对手是正在形成的IEEE802.16标准，称为WiMAX。这个标准承诺固定用户40Mbit/s和移动用户15Mbit/s的宽带无线接入，覆盖范围在数公里。规范的细节还在制定中。

601

D.3 无线短距离网络标准

最后这节归纳紫蜂、蓝牙以及UWB的主要特征，这些标准支持范围广泛的短距离无线网络应用。这些标准遵从IEEE 802.15标准，IEEE802.15标准是IEEE关于称为Wireless Personal Area Networks（WPANs）的短距离无线网络的协议族。蓝牙运行在2.4GHz的自由频段，紫蜂除了运行在2.4GHz自由频段外，还运行在800MHz以及900MHz自由频段，而UWB运行在一个广泛的频率范围内，在现存系统的频率范围下运行。紫蜂和蓝牙包含链路层、MAC层以及更高层次的规范，而UWB仅仅定义了链路层协议。表D-7归纳了紫蜂、蓝牙以及UWB的主要特征。

表D-7 短距离无线网络标准

	紫蜂（802.15.4）	蓝牙（802.15.1）	UWB（802.15.3）
频率范围（GHz）	2.4-2.4835	2.4-2.4835	3.1-10.6
带宽（MHz）	83.5	83.5	7500
调制	BPSK,OQPSK,DSSS	GFSK,FHSS	BPSK,QPSK,OFDM或DSSS
最大数据速率（Mbit/s）	0.25	1	100
覆盖范围（m）	30	10(100)	10
功耗（mW）	5~20	1(100)	80mW~150mW
接入	CSMA/CA（可选TD）	TD	未定义
组网	网状/星形/树形	子网集群（8节点）	未定义

紫蜂包含了遵从IEEE802.15.4标准的链路层和MAC层协议，以及更高层协议，涉及ad-hoc组网（网状，星形或树形），功率控制，以及安全性。紫蜂采用PSK调制以及DSSS支持高达250kbit/s的数据速率。紫蜂通常的应用目标要求相对低的数据速率，低负载以及大规模的网络。功率效率是其要点，其目标是节点依靠单次电池充电可以工作数月以至数年。

602

相对于紫蜂，蓝牙提供高达1Mbit/s的数据速率，包括三个有保证的低延迟话音信道，使用GFSK调制和FHSS。蓝牙通常使用1mW的功率传输，其范围为10m，通过增加功率到100mW，范围可以扩展到100m。网络由最多八个节点的子网集群构成，其中一个节点作为主节点，其他节点为从节点。TD用于信道接入，主节点调整FH序列，并同步从节点。当一个节点为多个子网络共有时可以构成扩展的网络或称散射网络。然而，由于FHSS的同步要求，通过这种方法形成大规模的网络是困难的。蓝牙的部分标准被IEEE正式接受，作为它的802.15.3标准。

相对紫蜂和蓝牙，UWB有明显的高数据速率，高达100Mbit/s。它也明显地占用更多的带宽，并且，为了防止干扰原来的频带用户，它有严格的功率限制。这样，它仅仅

适用短距离范围的室内应用。UWB仅仅定义了链路层技术，所以，它需要兼容MAC层协议和高层协议以成为无线网络标准。调制方式为BPSK或者QPSK，而其竞争阵营推荐OFDM或者DSSS叠加在调制上。UWB很可能成为IEEE 802.15.3标准的链路层技术，IEEE 802.15.3标准为支持图像和多媒体应用的无线网络的标准族。

参考文献

[1] T. S. Rappaport, *Wireless Communications – Principles and Practice*, 2nd ed., Prentice-Hall, Englewood Cliffs, NJ, 2001.

[2] J. D. Vriendt, P. Lainé, C. Lerouge, and X. Xu, "Mobile network evolution: A revolution on the move," *IEEE Commun. Mag.*, pp. 104–11, April 2002.

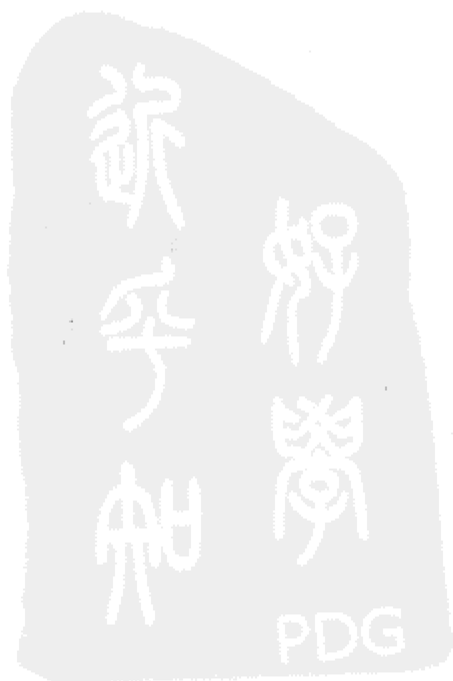
[3] D. Porcino and W. Hirt, "Ultra-wideband radio technology: Potential and challenges ahead," *IEEE Commun. Mag.*, pp. 66–74, July 2003.

[4] I. Poole, "What exactly is … ZigBee?" *IEEE Commun. Eng.*, pp. 44–5, August/September 2004.

[5] S. Haykin and M. Moher, *Modern Wireless Communications*, Prentice-Hall, Englewood Cliffs, NJ, 2005.

[6] W. Stallings, *Wireless Communications and Networks*, 2nd ed., Prentice-Hall, Englewood Cliffs, NJ, 2005.

603



索　　引

索引中的页码为英文原书页码，与本书中页边标注的页码一致。

D

E

F

G

H

I

K

L

M

N

O

P

Q

R

S

T

U

V

W

Z